Mathematical Institute
Slovak Academy of Sciences

# Geometry of families of states: from classical to quantum

Doktorská dizertačná práca

Anna Jenčová

Bratislava, január 2017

# Declaration

The thesis consists of works written in the years 2001-2016. Ten research papers were published in scientific journals, one is a chapter of a book and one was published in an IEEE conference proceedings. I hereby declare that none of these works were used before to obtain a scientific degree. A full list of the papers follows.

[IG1] A. Jenčová, Geometry of quantum states: dual connections and divergence functions, *Rep. Math. Phys.* **47** (2001), 121-138

[IG2] A. Jenčová, Generalized relative entropies as contrast functionals on density matrices, *Int. J. Theor. Phys.* **43**, (2004), 1635-1649

[IG3] A. Jenčová, Flat connections and Wigner-Yanase-Dyson metrics, *Rep. Math. Phys.* **52** (2003), 331–351

[IG4] A. Jenčová, Quantum information geometry and non-commutative $L_p$ spaces, *IDAQP* **8** (2005), 215–233

[IG5] A. Jenčová, A construction of a nonparametric quantum information manifold, *J. Funct. Anal.* **239** (2006), 1-20

[IG6] A. Jenčová, On quantum information manifolds, In: *Algebraic and Geometric Methods in Statistics*, Cambridge University Press 2010

[CE1] A. Jenčová, Quantum hypothesis testing and sufficient subalgebras, *Lett. Math. Phys.* **93** (2010), 15

[CE2] A. Jenčová, Reversibility conditions for quantum operations, *Rev. Math. Phys.* **24** (2012), 1250016

[CE3] A. Jenčová, Comparison of quantum binary experiments, *Rep. Math. Phys.* **70** (2012), 237-249

[CE4] A. Jenčová, Comparison of quantum channels and quantum statistical experiments, 2016 IEEE International Symposium on Information Theory (ISIT), 2249 - 2253, *IEEE Conference Publications*, 2016

[CS1] A. Jenčová, Generalized channels: Channels for convex subsets of the state space, *J. Math. Phys.* **53** (2012), 012201

[CS2] A. Jenčová, Base norms and discrimination of generalized quantum channels, *J. Math. Phys.* **55** (2013), 022201

Anna Jenčová, Bratislava, January 2017

# Preface

One of the fundamental features of quantum theory is its probabilistic nature. The theory provides predictions about probabilities rather than the events themselves and it is not possible to explain this indeterminism by lack of knowledge or presence of hidden variables. The classical probability theory cannot encompass the truly quantum properties. Mathematical description of quantum mechanics has to incorporate features like uncertainty principle, incompatibility of observables, superposition principle and entanglement. These features became powerful resources in quantum information science.

The difference between the classical and quantum state spaces is already well understood. While the classical state space is a Choquet simplex, quantum states have a more complicated structure. In the present thesis, we aim at the study of more specific properties of parametrized families of states. In the classical case, these properties are fundamental in theoretical statistics and asymptotic estimation theory, with applications also in other areas. The quantum case is often quite different and the properties have to be reformulated in a nontrivial way to recapture the classical results. The purpose of this work is to gain some understanding of the similarities and differences between the classical and quantum structures.

A parametrized family of states is called a statistical model, or a a statistical experiment. It represents a prior knowledge of the true state of some system, or the distribution from which some data are sampled. The states may be labeled by some interesting parameter and our ability to estimate this parameter depends on the geometry of the set of states and on the parametrization. This lead to introduction of a differential-geometrical structure on statistical models, studied by information geometry. In the framework of decision theory, the performance of available decision rules for various statistical tasks is studied and an ordering and a distance-like measure on statistical experiments can be defined by their comparison. Special families of states appear in quantum information theory, where quantum channels, or some more specific protocols, can be identified with certain convex subsets of a multipartite quantum state space. Statistical tasks, such as estimation or discrimination problems, appear naturally also in this context and the geometric structure of the state space plays a decisive role.

# Objectives of the thesis

The aim of this work is to find quantum versions of the results of two important theories, dealing with parametrized families of probability distributions and their structure: information geometry and theory of comparison of statistical experiments. As a tool for one of these tasks, but also as an interesting question in its own right, the convex structure of the set of quantum channels and its role in statistical decision theory is investigated. The particular problems solved in the thesis are the following:

- On the manifold of all positive definite complex matrices of a given dimension, we show that the condition of dual flatness singles out a unique family of dualistic structures with a monotone Riemannian metric.
- On the set of faithful states on a von Neumann algebra, we construct a Banach manifold structure, corresponding to the classical Pistone-Sempi construction, and investigate its behaviour under quantum channels.
- We investigate affine connections on the state space of a von Neumann algebra, obtained by embeddings into noncommutative $L_p$-spaces, their duality and the corresponding canonical divergences.
- We find conditions for sufficiency of a quantum channel with respect to a set of states, given in terms of some information-theoretical quantities such as error probabilities of hypothesis testing or quantum Fisher information.
- Different forms of Blackwell's informativity for quantum experiments are compared: informativity with respect to all decision problems and informativity with respect to testing problems.
- We find a fully quantum version of Le Cam's randomization criterion with a clear operational interpretation.
- Measurements and channels on convex subsets of the state space are studied, exploring their convex structure, the corresponding base norms and their relation to the tasks of statistical decision theory.

The thesis consists of twelve research papers, divided into three chapters according to their main subject. In Part I below, we give an introduction to each subject, a brief overview of the content of the corresponding works and a discussion of further research and open problems. The papers can be found in Part II.

# Contents

# Part I

# Summary of the results

# Chapter 1

# Basic definitions

The notion of a state is central to this work. For our purposes, it is a purely mathematical object. A *classical state* is a probability distribution on some measurable space $(\Omega, \mathcal{A})$, we denote by $\mathcal{S}(\Omega, \mathcal{A})$ the set of all states on $(\Omega, \mathcal{A})$. If $P \in \mathcal{S}(\Omega, \mathcal{A})$ is absolutely continuous with respect to some $\sigma$-finite measure $\mu$, it is represented by its density function

$$f \equiv \frac{dP}{d\mu} \in L_1(\Omega, \mathcal{A}, \mu)^+, \qquad \int_\Omega f d\mu = 1.$$

The set of all density functions with respect to $\mu$ will be denoted by $\mathcal{S}(\Omega, \mathcal{A}, \mu)$.

A *quantum state* is a normal positive unital functional on some von Neumann algebra $\mathcal{M}$ representing the observables of a quantum system. This definition contains classical states, in the case when $\mathcal{M}$ is abelian. The set of all states on $\mathcal{M}$ will be denoted by $\mathcal{S}(\mathcal{M})$. If $\mathcal{M} = B(\mathcal{H})$ is the algebra of bounded operators on a Hilbert space $\mathcal{H}$, states are represented by density operators, that is, positive trace class operators with unit trace. The set of all density operators on $\mathcal{H}$ will be denoted by $\mathcal{S}(\mathcal{H})$. We will often deal with finite dimensional Hilbert spaces, in which case the states can be identified with density matrices.

Transformations of states are represented by *stochastic maps*, also called *channels* or coarse-grainings. In the classical case, they can be defined simply as affine maps between state spaces. It is easy to see that such a map $\mathcal{S}(\Omega_1, \mathcal{A}_1, \mu_1) \to \mathcal{S}(\Omega_2, \mathcal{A}_2, \mu_2)$ extends to a positive map $L_1(\Omega_1, \mathcal{A}_1, \mu_1) \to L_1(\Omega_2, \mathcal{A}_2, \mu_2)$ preserving the norms of positive operators. In the quantum case, the channels are defined as preduals of unital normal completely positive maps between von Neumann algebras (but sometimes weaker positivity conditions are required). In finite dimensions, channels are identified with completely positive trace preserving maps between algebras of operators.

As special cases, we will encounter channels between quantum and classical state spaces. The quantum-to-classical channels are interpreted as *measurements*, assigning to each quantum state a corresponding probability distribution on the set of measurement outcomes. Any measurement can be uniquely represented by a positive operator valued measure (*POVM*), [38] and conversely any POVM defines a measurement. A POVM is a map $\mathcal{A} \ni A \mapsto M(A) \in \mathcal{M}^+$, which is $\sigma$-additive and normalized, $M(\Omega) = I$. We will only deal with the situation when $\Omega$

is a finite set, in which case any POVM is a collection of positive operators $M_i \in \mathcal{M}$ such that $\sum_i M_i = I$. Similarly, in this case, any classical-to-quantum channel, mapping classical states to quantum ones, can be identified with a finite set of quantum states, parametrized by elements of $\Omega$.

A *divergence* is an information-theoretic measure of difference of two states. Such a measure has to be a *contrast functional*, which means that it is nonnegative and equal to zero if and only if the states are equal. Another natural assumption is that it is nonincreasing under stochastic maps, since transformations of states cannot increase their distinguishability. For a divergence $D$, the inequality

$$D(T(\rho), T(\sigma)) \leq D(\rho, \sigma),$$

for all pairs of states $(\sigma, \rho)$ and all channels $T$ is referred to as the *data processing inequality*.

An essential example of a classical divergence measure is the *relative entropy* [48], also called *Kullback-Leibler divergence* or *$I$-divergence*. For two probability distributions $P \ll Q$, it is defined as

$$S(P\|Q) = \int \log(\frac{dP}{dQ})dP.$$

More generally, for any convex function $f$ on $\mathbb{R}^+$, the *$f$-divergence* [18, 54] is defined as

$$S_f(P\|Q) = \int f(\frac{dP}{dQ})dQ.$$

Note that relative entropy is obtained for $f = x\log(x)$. Another special case is the *$\alpha$-divergence*

$$S_\alpha(P\|Q) = \frac{4}{1-\alpha^2}\left(1 - \int (\frac{dP}{dQ})^{\frac{1-\alpha}{2}} dQ\right), \qquad \alpha \neq \pm 1.$$

For density operators, the *Umegaki relative entropy* [82] has the form

$$S(\rho\|\sigma) = \operatorname{Tr}\rho(\log(\sigma) - \log(\rho)),$$

if the support of $\rho$ is included in the support of $\sigma$ and is infinite otherwise. In the setting of von Neumann algebras, quantum relative entropy was defined by Araki [4] using the relative modular operator. Quantum versions of $f$-divergences, also called *quasi-entropies*, were introduced by Petz [63, 64]. For density matrices (with some conditions on their supports), these have the form

$$S_f(\rho\|\sigma) = \operatorname{Tr}\sigma^{1/2} f(\Delta_{\rho,\sigma})(\sigma^{1/2}),$$

where $\Delta_{\rho,\sigma}$ is a positive operator on the Hilbert space of matrices equipped with the Hilbert-Schmidt inner product $(X, Y) = \operatorname{Tr} X^*Y$, defined as $\Delta_{\rho,\sigma} : X \mapsto \rho X \sigma^{-1}$. The quasi-entropies satisfy data processing inequality if the function $f$ is operator convex. In particular, the version of the $\alpha$-divergence for density operators is given by

$$S_\alpha(\rho\|\sigma) = \frac{4}{1-\alpha^2}\left(1 - \operatorname{Tr}\rho^{\frac{1-\alpha}{2}}\sigma^{\frac{1+\alpha}{2}}\right),$$

but the requirement of monotonicity restricts the values of $\alpha$ to the interval $[-3, 3]$. The relative entropy and $\alpha$-divergences (or the closely related Rényi relative entropies), both classical and quantum, are important distinguishability measures in information theory and statistics.

# Chapter 2

# The structures of information geometry

The aim of classical information geometry is the study of differential geometrical structures derived from the properties of statistical models. These structures are already well understood and the theory has a number of important applications e.g. in asymptotic estimation theory, information theory, machine learning, statistical mechanics, biology and theory of neural networks. Interested readers may refer to the monograph [3] by Amari and Nagaoka.

## 2.1 Information geometry for parametric models

A classical statistical model is a parametrized family

$$\mathcal{P} = \{p_\theta,\ \theta \in \Theta\}, \qquad p_\theta \in \mathcal{S}(\Omega, \mathcal{A}, \mu).$$

If $\Theta \subseteq \mathbb{R}^n$ is an open set and the parametrization $\theta \mapsto p_\theta$ is sufficiently regular, it introduces a differentiable manifold structure in $\mathcal{P}$. It was first observed in the works by Rao [71] and Jeffreys [39] that such a manifold can be endowed with a Riemannian metric $\lambda^F$, given by the Fisher information

$$\lambda^F_{ij}(\theta) = E_{p_\theta}[\partial_i \log(p_\theta)\partial_j \log(p_\theta)], \qquad \partial_i := \frac{\partial}{\partial\theta_i}.$$

The well-known Cramér-Rao inequality for the variance of unbiased estimators shows that this metric expresses how precisely a point on the manifold can be distinguished from other points in its neighborhood using statistical methods.

As it turned out, a Riemannian structure is not enough to capture the statistical properties of the model. The importance of exponential families was pointed out by Efron [21], who defined the statistical curvature for 1-dimensional models and shown its role in asymptotic estimation theory. Based on this observation, Dawid [20] introduced the exponential affine connection $\nabla^{(e)}$ on the manifold and proved that the statistical curvature is precisely the embedding curvature of the model with respect to $\nabla^{(e)}$. Amari [1] extended this work to a family of $\alpha$-connections parametrized by $\alpha \in \mathbb{R}$, containing $\nabla^{(e)}$ for the value $\alpha = 1$. These connections are defined by a pullback of the natural affine structure on the set of measurable functions via the *Amari embeddings*

$$p_\theta \mapsto g_\alpha(p_\theta), \qquad g_\alpha(x) = \begin{cases} \frac{2}{1-\alpha} x^{\frac{1-\alpha}{2}}, & \alpha \neq 1 \\ \\ \log(x), & \alpha = 1. \end{cases} \tag{2.1}$$

Another special element in this family is the mixture connection $\nabla^{(m)} := \nabla^{(-1)}$, arising from the convex structure of the state space. The same class of connections is obtained as an affine mixture

$$\nabla^{(\alpha)} = \frac{1-\alpha}{2}\nabla^{(m)} + \frac{1+\alpha}{2}\nabla^{(e)}, \qquad \alpha \in \mathbb{R}.$$

These geometric structures on a statistical model can be also introduced using $f$-divergences [22]. If $f$ is normalized such that $f''(1) = 1$, then

$$(\lambda^F)_{ij}(\theta) \equiv \lambda^F(\partial_i, \partial_j)|_\theta = \partial_i \partial'_j S_f(p_\theta \| p_{\theta'})|_{\theta=\theta'}, \tag{2.2}$$

$$\Gamma^{(\alpha)}_{ijk}(\theta) \equiv \lambda^F(\nabla^{(\alpha)}_{\partial_i}\partial_j, \partial_k)|_\theta = \partial_i \partial_j \partial'_k S_f(p_\theta \| p_{\theta'})|_{\theta=\theta'}, \tag{2.3}$$

with $\partial'_i := \frac{\partial}{\partial \theta'_i}$ and $\alpha = 2f'''(1) + 3$.

An important feature of the family of $\alpha$-connections is that $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ are dual with respect to the Fisher metric $\lambda^F$. Manifolds with a dualistic structure $(\lambda, \nabla, \nabla^*)$, consisting of a Riemannian metric and a pair of dual connections, were investigated by Nagaoka and Amari [62]. One of their deep results is that if the manifold is dually flat, a pair of dual coordinate systems exists, connected by a strictly convex potential function $\Phi$. The corresponding Bregman divergence defines a distance-like measure on the manifold, called a *canonical divergence*:

$$D(p_{\theta_1}, p_{\theta_2}) = D_\Phi(\theta_1, \theta_2) := \Phi(\theta_1) - \Phi(\theta_2) - \partial\Phi(\theta_2)(\theta_1 - \theta_2).$$

For statistical manifolds, the relative entropy and the family of $\alpha$-divergences can be derived in this way. By construction, the canonical divergence satisfies a generalized Pythagorean relation and certain projection theorems hold, which are important in optimization tasks.

A simple example of a flat statistical manifold is $\mathcal{P}_n$, consisting of strictly positive probability measures on $n$ points. It is easy to see that $\mathcal{P}_n$ is $\nabla^{(\pm 1)}$-flat. For $\alpha \neq \pm 1$, the $\alpha$-divergences are obtained by restriction from the extended manifold $\hat{\mathcal{P}}_n$ of strictly positive functions, which is $\nabla^{(\alpha)}$-flat for all $\alpha \in \mathbb{R}$.

The family of connections $\{\nabla^{(\alpha)},\ \alpha \in \mathbb{R}\}$ was obtained independently by Cencov [12] in a quite different approach. Cencov introduced a category of statistical models with stochastic maps as morphisms and investigated geometric structures that are invariant with respect to isomorphisms in this category. He proved that up to multiplication by a scalar, $\lambda^F$ is the unique Riemannian metric and $\nabla^{(\alpha)}$, $\alpha \in \mathbb{R}$ are the only affine connections with this property.

## 2.2 Nonparametric information geometry

Nonparametric information geometry was introduced by Pistone and Sempi [69] and further developed in [68]. Here the model consists of all elements in $\mathcal{S}(\Omega, \mathcal{A}, \mu)$ equivalent to a given one. This manifold has to be modelled on a Banach space, but the embeddings into $L_p$ spaces, used to define the connections in the parametric case, are not suitable for this, simply because the positive cone in these spaces has an empty interior. Pistone a Sempi used the *exponential Orlicz space* $L_\Phi(p)$, given by the Young function

$$\Phi(x) = \cosh(x) - 1.$$

The subspace of centered functions in $L_\Phi(p)$ then parametrizes the neighborhood of $p$ on the manifold of probability distributions equivalent to $p$ as an exponential family. The exponential Orlicz space is not even reflexive, so that the geometric structures introduced for parametric models have no straightforward generalization. More precisely, there is no Riemannian structure and the Fisher information can be introduced as a continuous bilinear functional, defined by differentiation of the cumulant generating functional. Affine connections on this manifold were studied in [24, 25]. In this case, the $\alpha$-connections live on separate fiber bundles and their duality corresponds to the Banach space duality of Orlicz spaces.

## 2.3 Geometry of quantum states: the finite dimensional case

The aim of quantum information geometry is the extension of the results of the classical theory to families of quantum states. In the simplest case, the quantum system is represented on an $n$-dimensional Hilbert space. It is then enough to study the geometry of the manifold $\mathcal{D}_n$ of positive definite density matrices of dimension $n$, since any model of sufficient regularity can be embedded into it. As an open subset of a finite dimensional real vector space, $\mathcal{D}_n$ has a natural affine and manifold structure. But already in this simple case, we encounter problems that do not appear in classical models.

The tangent space $T_\rho(\mathcal{D}_n)$ at $\rho \in \mathcal{D}_n$ is isomorphic to the space of traceless Hermitian $n \times n$ matrices. With this identification, any Riemannian metric on $\mathcal{D}_n$ has the form

$$\lambda_\rho(X, Y) = \operatorname{Tr} X J_\rho(Y), \qquad X, Y \in T_\rho(\mathcal{D}_n), \tag{2.4}$$

where $J_\rho$ is a suitable operator on matrices. An important example is the *symmetric logarithmic derivative* $J_\rho^{SLD}$, defined by

$$J_\rho^{SLD}(Y) = H, \qquad 2Y = \rho H + H\rho.$$

This choice defines the *SLD-metric* $\lambda^{SLD}$ [38], which is usually considered as the quantum Fisher information, since it satisfies an analog of the classical Cramér-Rao inequality. But, in contrast to the classical case, this inequality is typically not optimal and cannot be attained even asymptotically.

Inspired by the uniqueness result in [12], Cencov and Morozova [13] studied Riemannian metrics on $\mathcal{D}_n$ which are nonincreasing under quantum channels. Such metrics are called *monotone*. It was proved that, unlike the classical case, there is a large family of such metrics. Later, it was shown by Petz [67] that a Riemannian metric on $\mathcal{D}_n$ is monotone if and only if the corresponding operator has the form

$$J_\rho^f := R_\rho^{-1} f(L_\rho R_\rho^{-1})^{-1}, \quad L_\rho(X) = \rho X,\ R_\rho(X) = X\rho \tag{2.5}$$

for some operator monotone function $f : (0,\infty) \to (0,\infty)$, which is *symmetric*, that is, $f(t) = tf(t^{-1})$ for all $t > 0$. With the normalization $f(1) = 1$, such a metric is called a *quantum Fisher information*.

The family of quantum Fisher informations contains the SLD-metric as the smallest element. Other important examples are given by the family of Wigner-Yanase-Dyson (*WYD*-) metrics $\lambda^\alpha (= \lambda^{-\alpha})$, parametrized by $\alpha \in [-3, 3]$ [32]. For $\alpha = \pm 1$, we obtain the Bogoljubov-Kubo-Mori (*BKM*-) metric, which is given as an infinitesimal version of the quantum relative entropy. The choice $\alpha = \pm 3$ yields the *RLD-metric*, which is the largest quantum Fisher information. The SLD-metric is not contained in the WYD-family. In another approach, Lesniewski and Ruskai [53] proved that any quantum Fisher information is obtained from a quasi-entropy as in (2.2).

Classical constructions of the $\alpha$-connections can be applied also in $\mathcal{D}_n$. In particular, the Amari embeddings

$$\rho \mapsto g_\alpha(\rho), \qquad \alpha \in \mathbb{R}$$

with $g_\alpha$ as in (2.1) can be used to pull back the affine structure of Hermitian matrices. As in the classical case, we will denote these connections by $\nabla^{(\alpha)}$, $\alpha \in \mathbb{R}$, and we put $\nabla^{(e)} := \nabla^{(1)}$ and $\nabla^{(m)} := \nabla^{(-1)}$. It is easy to see that both $\nabla^{(e)}$ and $\nabla^{(m)}$ are flat, however, they are not dual with respect to the usual SLD-metric. Moreover, it was proved by Nagaoka that the dual to $\nabla^{(m)}$ with respect to a monotone metric $\lambda$ is not torsion-free, hence not flat, unless $\lambda = \lambda^{BKM}$ [61]. As in the classical case, $\nabla^{(\alpha)}$ is not flat for $\alpha \neq \pm 1$, but the extension to the extended manifold $\hat{\mathcal{D}}_n$ of positive definite matrices is flat for all $\alpha$. This extension will be also denoted by $\nabla^{(\alpha)}$. Duality of $\nabla^{(\pm\alpha)}$ with respect to a quantum Fisher information was studied also in [31, 25, 27].

## 2.4 Nonparametric quantum information manifolds

The main obstacle in the construction of a nonparametric (infinite dimensional) quantum information manifold was the lack of a suitable non-commutative counterpart of an Orlicz space. For sets of density operators on a separable Hilbert space, some constructions were proposed in [79, 26] using small perturbations of the Hamiltonian at each point of the manifold, or by a quantum Young function [78]. In [23], the $\alpha$-connections for $\alpha \in (-1,1)$ were defined on manifolds of states on a semifinite von Neumann algebra by Amari embeddings, which map the density operators into the non-commutative $L_p$-space $L_p(\mathcal{M},\tau)$, $p = \frac{2}{1-\alpha}$. It was shown that duality of the connections is in fact obtained from duality of the spaces $L_{\frac{2}{1-\alpha}}$ and $L_{\frac{2}{1+\alpha}}$ and it was pointed out that uniform convexity of $L_p$-spaces is crucial for projection of the $L_p$-space geometry onto the set of states.

## 2.5 The results

In the first three papers in this chapter, constructions of dually flat affine connections on $\mathcal{D}_n$ and $\hat{\mathcal{D}}_n$ are discussed. In the rest, properties of noncommutative $L_p$ spaces and the quantum (Araki) relative entropy are used for a construction of a Banach manifold structure on the set of normal states of a von Neumann algebra.

### 2.5.1 Content of the papers [IG1–6]

**[IG1]** On the extended manifold $\hat{\mathcal{D}}_n$ with a monotone Riemannian metric $\lambda$, we study dualistic structures $(\lambda, \nabla^{(\alpha)}, \nabla^{(\alpha)^*})$, where $\nabla^{(\alpha)^*}$ is defined as the dual connection to $\nabla^{(\alpha)}$ with respect to $\lambda$. We compute the torsion of these connections and it is pointed out that the dual connections $\nabla^{(\alpha)^*}$ are in general not torsion free. If $\lambda = \lambda^\alpha$ for some $\alpha \in [-3,3]$, we obtain the dually flat structure $(\lambda^\alpha, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$.
The dualistic structure is then projected onto the manifold of density matrices $\mathcal{D}_n$ and the embedding curvature and Riemannian curvature tensors are computed. We also define divergence functions on $\hat{\mathcal{D}}_n$, considering the $\nabla^{(\alpha)*}$-geodesics connecting two points and using the fact that one-dimensional submanifolds are always torsion-free. For $\alpha = \pm 1$, we do the same also for $\mathcal{D}_n$. As examples, we obtain the quantum $\alpha$-divergences and the Umegaki and Belavkin-Staszevski [5] versions of the relative entropy.

**[IG2]** Here we investigate a different construction of a dualistic structure. Using the classical results by Eguchi, see (2.2), (2.3), and the work by Lesniewski and Ruskai [53], we obtain a monotone metric and a dual pair of torsion-free affine connections in $\hat{\mathcal{D}}_n$ from a quasi-entropy. These connections coincide with some $\nabla^{(\alpha)}$ on $\hat{\mathcal{P}}_n$, but the parameter $\alpha$ is restricted to the interval $[-3,3]$. Using the theory of statistical manifolds devised by Lauritzen [50], we compute the Riemannian curvature tensor. In particular, using Umegaki relative entropy, we obtain the BKM-metric and the family of connections

$$\frac{1-\alpha}{2}\nabla^{(m)} + \frac{1+\alpha}{2}\nabla^{(e)}, \qquad \alpha \in [-1,1].$$

It is pointed out that the Riemannian curvature of these connections cannot be 0 unless $\alpha = \pm 1$, so that these connections must be different from $\nabla^{(\alpha)}$ if $\alpha \neq \pm 1$.

**[IG3]** This paper finishes both previous works by proving that in either construction, the unique dually flat structures with respect to a monotone metric are given by $(\lambda_\alpha, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$, $\alpha \in [-3, 3]$. For this, we apply the results of Lauritzen and some tools of matrix analysis.

**[IG4]** We use the natural bijective mapping of the predual of a von Neumann algebra onto the noncommutative $L_p$-space on $\mathcal{M}$ with respect to a faithful normal semifinite weight, defined by Masuda [56]. Using duality and uniform convexity of the $L_p$-spaces, we obtain a pair of coordinate systems on the set of positive normal functionals, connected by norm-continuous maps that are uniformly continuous when restricted to states. The coordinate systems are related by Legendre transforms. Uniform Fréchet differentiability of the $L_p$-norms allows us to define a divergence function, which turns out to be the quantum $\alpha$-divergence $S_\alpha$, $\alpha = \frac{p-2}{p}$. A generalized Pythagorean relation and projection theorems for $S_\alpha$ are derived.

**[IG5]** We define a version of the exponential Orlicz space with respect to a faithful normal state $\varphi$ on $\mathcal{M}$, using a Young function constructed from the convex conjugate of the Araki relative entropy

$$c_\varphi(h) = \sup_{\omega\in\mathcal{S}(\mathcal{M})} \omega(h) - S(\omega\|\varphi), \qquad h = h^* \in \mathcal{M}.$$

The quantum exponential Orlicz space with respect to $\varphi$ is the completion of the space of self-adjoint elements in $\mathcal{M}$ with respect to the corresponding Orlicz norm. It is proved that the dual space is generated by positive normal functionals, such that the relative entropy with respect to $\varphi$ is finite.
By the relative entropy approach to perturbation of states, the set of faithful normal states is endowed with a manifold structure modeled on the subspace of centered elements in this space, together with exponential and mixture connections, living on the tangent and cotangent space, respectively.

**[IG6]** Here we provide another construction of an exponential Orlicz space as the space of continuous affine functions on the compact convex set

$$\mathcal{S}_{1,\varphi} = \{\omega \in \mathcal{S}(\mathcal{M}),\ S(\omega\|\varphi) \le 1\}.$$

We prove that this construction is equivalent to the previous one. It is shown that channels between von Neumann algebras extend to morphisms of Banach manifolds and this construction is functorial. Further, for a pair of faithful states $\psi, \varphi \in \mathcal{S}(\mathcal{M})$ contained in the same connected component of the manifold, the adjoint of the channel extends to a mapping of the corresponding coordinates if and only if the channel is sufficient (or reversible) with respect to $\{\psi, \varphi\}$ (see Section 3.4).

### 2.5.2 Conclusions and open problems

In the classical case, there is a unique 1-parameter class of dualistic structures, satisfying the Cencov invariance condition [12]. Moreover, these structures are all flat on the simplest classical manifold $\hat{\mathcal{P}}_n$. As we have seen, there is a variety of such structures in the quantum case and classical constructions lead to different results. But the requirement that the structures should be flat on the simplest quantum manifold $\hat{\mathcal{D}}_n$ singles out the unique 1-parameter class $(\lambda^\alpha, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$, where, in contrast to the classical case, all the metrics are different and the value of the parameter is restricted to the interval $[-3, 3]$.

We constructed a Banach manifold structure on the set of faithful states of a von Neumann algebra $\mathcal{M}$, derived from Araki relative entropy. Channels induce contractions on the corresponding tangent spaces and the coordinates are preserved on subsets of states if and only if the channel is sufficient with respect to them. We also proved that the quantum $\alpha$-divergences appear as canonical divergences for dually flat connections, obtained by embeddings in non-commutative $L_p$-spaces.

The relation of the constructed Orlicz spaces to known constructions of non-commutative $L_p$-spaces is not known. Related questions are whether the constructed $\alpha$-embeddings are compatible with the exponential manifold structure, differentiability of the divergences and possibility to introduce some form of Fisher information on the manifold. It was also pointed out in [IG6] that the proper quantum counterpart of the exponential Orlicz space is not the constructed space $B_\varphi$, but rather its second dual.

After our works on the nonparametric version were published, a new definition of a quantum Orlicz space was proposed by Labuschagne [49]. Based on this definition, a nonparametric quantum information manifold was constructed by Labuschagne and Majewski [55], with an interpretation in description of large regular statistical systems, both classical and quantum. It is an interesting question how this is related to our construction.

For classical information manifolds, it was proved by Amari [2] that the classical $\alpha$-divergences are the unique $f$-divergences that are Bregman divergences at the same time. It seems plausible that this is the case also for the quantum $\alpha$-divergences, but a rigorous proof is left for future work.

### 2.5.3 Further related works by the author

- A. Jenčová, Dualistic properties of the manifold of quantum states, In: *Disordered and complex systems*, AIP Conference Proceedings, Melville, New York 2001
- A. Jenčová, Quantum information geometry and standard purification, *J. Math. Phys.*, 43 (2002), 2187-2201
- A. Jenčová, Geodesic distances on density matrices, *J. Math. Phys.* 45 (2004), 1787-1794

# Chapter 3

# Comparison of channels and statistical experiments

## 3.1 Comparison of classical statistical experiments

In statistical decision theory, a *statistical experiment* is defined as a triple $\mathcal{E} = (\Omega, \mathcal{A}, \mathcal{P})$, where $\mathcal{P} = \{P_\theta,\ \theta \in \Theta\} \subseteq \mathcal{S}(\Omega, \mathcal{A})$ and $\Theta$ is a parameter set. In most cases, it is assumed that the experiment is *dominated*, which means that we may assume that $\mathcal{P} \subseteq \mathcal{S}(\Omega, \mathcal{A}, \mu)$ for some $\sigma$-finite measure $\mu$. If $\Theta = \{\theta_1, \theta_2\}$, the experiment is called *binary*.

The set $\mathcal{P}$ is interpreted as the set of possible probability distributions from which data are sampled. Based on the sample, a decision $d$ is chosen from the set of decisions $D$. We will work mostly with the situation when $D$ is a finite set, but in general, $D$ is a topological space. The decision is evaluated according to a *loss function* $W : \Theta \times D \to \mathbb{R}$, where it is assumed that the functions $W_\theta := W(\theta, \cdot)$ are continuous and bounded, with $\|W_\theta\| = \sup_{t\in D} |W_\theta(t)|$. The triple $(\Theta, D, W)$ is called a *decision problem*. If $D$ consists of two elements, then $(\Theta, D, W)$ is called a *testing* problem.

A most general strategy for decision, or a *decision rule*, is given by a stochastic map $M : \mathcal{S}(\Omega, \mathcal{A}, \mu) \to \mathcal{S}(D, \mathcal{B}_0(D))$, where $\mathcal{B}_0(D)$ is the Baire field over $D$. The *risk* for the given value of the parameter is computed as

$$R_{\mathcal{E},D,W,\theta}(M) = \int_D W_\theta \, dM(P_\theta).$$

The set of all decision rules will be denoted by $\mathcal{R}(\mathcal{E}, D)$.

Let $\mathcal{E} = (\Omega_1, \mathcal{A}_1, \mathcal{P})$ and $\mathcal{F} = (\Omega_2, \mathcal{A}_2, \mathcal{Q})$ be dominated experiments with the same parameter set $\Theta$ and dominating measures $\mu_\mathcal{E}$ and $\mu_\mathcal{F}$. The following preorder on statistical experiments was introduced by Blackwell [6]: the experiment $\mathcal{E}$ is *more informative* than $\mathcal{F}$, in notation $\mathcal{E} \succeq \mathcal{F}$, if for any decision problem $(\Theta, D, W)$ and any $M \in \mathcal{R}(\mathcal{F}, D)$ there is some $N \in \mathcal{R}(\mathcal{E}, D)$, such that

$$R_{\mathcal{E},D,W,\theta}(N) \leq R_{\mathcal{F},D,W,\theta}(M), \qquad \forall \theta \in \Theta.$$

If $T : \mathcal{S}(\Omega_1, \mathcal{A}_1, \mu_\mathcal{E}) \to \mathcal{S}(\Omega_2, \mathcal{A}_2, \mu_\mathcal{F})$ is a channel such that $\mathcal{Q} = T(\mathcal{P})$, we say that $\mathcal{F} =: T(\mathcal{E})$ is a *randomization* of $\mathcal{E}$. If $(\Theta, D, W)$ is any decision problem and $M \in \mathcal{R}(\mathcal{F}, D)$, then $M \circ T \in \mathcal{R}(\mathcal{E}, D)$ and we have

$$R_{\mathcal{F},D,W,\theta}(M) = \int_D W_\theta \, dM(T(P_\theta)) = \int_D W_\theta \, d(M \circ T(P_\theta)) = R_{\mathcal{E},D,W,\theta}(M \circ T).$$

It follows that $\mathcal{E} \succeq T(\mathcal{E})$. The following theorem is one of the basic results of the theory of statistical experiments.

**Theorem 1** (Blackwell-Sherman-Stein (BSS) [6, 73, 76])**.** *$\mathcal{E}$ is more informative than $\mathcal{F}$ if and only if $\mathcal{F}$ is a randomization of $\mathcal{E}$.*

Le Cam [51] extended the above preorder on statistical experiments as follows. He defined the *deficiency* of $\mathcal{E}$ with respect to $\mathcal{F}$ as

$$\delta(\mathcal{E}, \mathcal{F}) := \inf_T \sup_\theta \|p_\theta - T(q_\theta)\|_1,$$

where the infimum is taken over all suitable channels. The *Le Cam distance*

$$\Delta(\mathcal{E}, \mathcal{F}) := \max\{\delta(\mathcal{E}, \mathcal{F}), \delta(\mathcal{F}, \mathcal{E})\}$$

is a pseudo-distance on the set of experiments with the same parameter set. Convergence with respect to this distance is important in the theory of asymptotic statistics, [52]. The following result is the celebrated Le Cam's randomization criterion.

**Theorem 2** (Le Cam's randomization criterion [51])**.** *Let $\mathcal{E}$, $\mathcal{F}$ be statistical experiments, $\epsilon \geq 0$. Then $\delta(\mathcal{E}, \mathcal{F}) \leq \epsilon$ if and only if for every $(D, W)$ and $M \in \mathcal{R}(\mathcal{F}, D)$ there is a decision rule $N \in \mathcal{R}(\mathcal{E}, D)$ such that*

$$R_{\mathcal{E},D,W,\theta}(M') \leq R_{\mathcal{F},D,W,\theta}(M) + \epsilon/2\|W_\theta\|.$$

Other variants of deficiency can be obtained by restricting to some special type of decision problems. In particular, deficiency with respect to testing problems is denoted by $\delta_2$ and the corresponding preorder by $\preceq_2$. In general, $\preceq_2$ is weaker than $\preceq$, but for binary experiments the two preorders are equivalent, [81], see also [77].

## 3.2 Comparison of quantum statistical experiments

A *quantum statistical experiment* is defined as a pair $\mathcal{E} = (\mathcal{M}, \mathcal{P})$, where, most generally, $\mathcal{M}$ is a von Neumann algebra and $\mathcal{P} \subseteq \mathcal{S}(\mathcal{M})$. This notion clearly contains (dominated) classical experiments, which correspond to quantum experiments on a commutative von Neumann algebra $\mathcal{M} = L_\infty(\Omega, \mathcal{A}, \mu_\mathcal{E})$. Let us denote the set of all quantum statistical experiments with parameter set $\Theta$ by $\mathcal{E}(\Theta)$.

Let $\mathcal{E} \in \mathcal{E}(\Theta)$ and let $(\Theta, D, W)$ be a decision problem. The decision rules in $\mathcal{R}(\mathcal{E}, D)$ are measurements (POVMs) with values in $(D, \mathcal{B}_0(D))$. In this work, we will consider only experiments on the algebra $B(\mathcal{H})$ for a finite dimensional Hilbert space $\mathcal{H}$. In this case, it is enough to consider decision problems with a finite decision set $D$. For $M \in \mathcal{R}(\mathcal{E}, D)$, the risk is computed as

$$R_{\mathcal{E},D,W,\theta}(M) = \sum_{d\in D} W_\theta(d) \mathrm{Tr}\, \rho_\theta M_d.$$

*Example* 3. A special type of a decision problem is multiple hypothesis testing. In this case, $\Theta = D = \{1, \dots, m\}$ and $W(i,j) = 1 - \delta_{ij}$. The task is to determine which of a given set of states $\{\rho_1, \dots, \rho_m\} \subset \mathcal{S}(\mathcal{H})$ is the true state of the system. Decision rules are POVMs $M = \{M_1, \dots, M_m\}$ and $\mathrm{Tr}\, \rho_i M_j$ is interpreted as the probability of choosing $\rho_j$if the true state is $\rho_i$. If prior probabilities $\lambda_1, \dots, \lambda_m$ are given, we are looking for a POVM with optimal Bayes risk, or equivalently with the maximum Bayes probability of success. Let $E$ be the *ensemble* $E := \{\lambda_i, \rho_i\}_{i=1}^m$. The maximum Bayes success probability is defined as

$$P_{succ}(E) := \max_M \sum_i \lambda_i \mathrm{Tr}\, \rho_i M_i, \tag{3.1}$$

where the maximization is over all POVMs with $m$ elements. Conditions for optimal POVMs were given in [36, 37], an explicit expression is not known in general.

*Example* 4. If we put $\Theta = D = \{0, 1\}$ in the previous example, we obtain the *discrimination problem* for two states $\rho_0$, $\rho_1$. Decision rules are given by $\{M, I - M\}$ for operators $0 \leq M \leq I$ in $B(\mathcal{H})$ and the maximal Bayes probability of success can be computed as [35, 38]

$$P_{succ}(E) = \Pi_\lambda(\rho_0, \rho_1) := \frac{1}{2}(1 - \|\lambda\rho_0 - (1-\lambda)\rho_1\|_1),$$

where $\|\rho\|_1 = \mathrm{Tr}\, |\rho|$ is the trace norm on $B(\mathcal{H})$. The optimal decision rule is $\{P, I - P\}$, where $P$ is the projection onto the support of the positive part of the operator $\lambda\rho_0 - (1-\lambda)\rho_1$.

The Blackwell preorder and deficiency can be extended to all experiments in $\mathcal{E}(\Theta)$ in an obvious way. If the experiment $\mathcal{F}$ is classical, the BSS theorem (Theorem 1) and Le Cam's randomization criterion (Theorem 2) hold [57, 40], but for arbitrary quantum experiments this is no longer true [58].

A fully quantum version of the BSS theorem was first obtained by Shmaya [75] and Buscemi [7]. In these works, either additional entanglement or composition of the experiment with a complete set of states is required. A quantum version of the randomization criterion was proved by Matsumoto [57], but the criterion is formulated in terms of quantum decision problems. This generalization of classical decision problems is natural from a mathematical point of view, but its operational significance is unclear.

## 3.3 Comparison of channels

Comparison of statistical experiments is closely related to comparison of channels. For two channels $T$ and $S$ with the same input space, we can say that $T$ is *less noisy* than $S$ if $S$ is a *post-processing* of $T$, that is, there is a channel $R$ such that $S = R \circ T$. Assume that the input space is a classical state space $\mathcal{S}(\Omega, \mathcal{A})$ with a finite set $\Omega$, then the channels can be interpreted as statistical experiments parametrized by the elements of $\Omega$. In this interpretation, randomization is the same as post-processing and the preorder $\preceq$ can be reformulated by comparing the optimal success probabilities in multiple hypothesis testing (see Example 3) for the ensembles obtained by applying $S$ and $T$ to any input ensemble.

*Remark* 5. For classical channels, a related ordering was introduced in a work by Shannon [72], where error probabilities of channel coding schemes are compared. This ordering is characterized by existence of channels $P_i$, $Q_i$ and probabilities $\lambda_i$, $i = 1, \dots, k$ such that $S = \sum_i \lambda_i P_i \circ T \circ Q_i$. Other orderings of classical channels can be found in [46, 19, 70, 9].

For classical channels, we obtain an obvious version of the BSS theorem and randomization criterion. However, as noted in the previous section, the BSS theorem in the classical formulation does not hold for quantum channels even if the common input space is classical. A stronger ordering is obtained if we consider ensembles on the input space coupled with an ancilla. As it turns out, with this ordering, the BSS theorem can be recovered. This remarkable result was first obtained by Chefles in [14] and was extended and refined in [7], in particular, it was proved that no entanglement in the input ensemble is needed. Some applications were already found in [11, 8, 10, 9].

## 3.4 Classical and quantum sufficient statistics

Let $T$ be a channel and let $\mathcal{E} \in \mathcal{E}(\Theta)$ be a statistical experiment in the input space of $T$. Let $T(\mathcal{E})$ be the corresponding randomization $\mathcal{E}$. The channel $T$ is called *sufficient with respect to* $\mathcal{E}$ if also $\mathcal{E}$ is a randomization of $T(\mathcal{E})$, that is, there is some channel $S$ such that

$$S \circ T(\rho_\theta) = \rho_\theta, \qquad \theta \in \Theta. \tag{3.2}$$

If $\mathcal{E}$ is classical, then by Theorem 1 and the paragraph below Example 4, such a channel exists if and only if $T(\mathcal{E}) \succeq \mathcal{E}$, so that $\mathcal{E}$ a $T(\mathcal{E})$ are equivalent in the sense of Blackwell's preorder. In fact, it is enough that we have $T(\mathcal{E}) \succeq_2 \mathcal{E}$, since the equivalence classes with respect to $\succeq$ and $\succeq_2$ are the same, [77]. For quantum experiments, validity of the corresponding statements is not known. A sufficient channel is also called a *statistical isomorphism* [77].

A special case of a sufficient classical channel is a *sufficient statistic*. This is a measurable map $f : (\Omega_1, \mathcal{A}_1) \to (\Omega_2, \mathcal{A}_2)$ such that the conditional probabilities $P_\theta[A|f] =: P[A|f]$,

$A \in \mathcal{A}_1$ do not depend from $\theta$. Let $T_f$ denote the channel $P \mapsto P^f$ then

$$S : Q \mapsto S(Q), \quad S(Q)(A) = \int P[A|f]dQ, \qquad A \in \mathcal{A}_1$$

is a channel satisfying $S \circ T_f(P_\theta) = S(P_\theta^f) = P_\theta$. This means that $T^f$ is sufficient with respect to $\mathcal{E}$. If $p_\theta = \frac{dP_\theta}{d\mu}$, sufficient statistics are characterized by the *factorization criterion*

$$p_\theta(\omega) = h(\omega)q_\theta(f(\omega)), \qquad \mu - a.e.,$$

where $q_\theta$ and $h$ are nonnegative measurable functions. This means that $p_\theta$ depends on $\theta$ only through $f$. If there is some $\theta_0$ such that $S(P_\theta \| P_{\theta_0})$ is finite for all $\theta$, then $f$ is sufficient with respect to $\mathcal{E}$ if and only if [48]

$$S(P_\theta^f \| P_{\theta_0}^f) = S(P_\theta, P_{\theta_0}).$$

This statement holds for all classical channels and all $f$-divergences with a strictly convex function $f$ [18, 54].

Sufficiency of quantum channels in the general form given by (3.2) was first investigated in the works by Petz, [65, 66], who studied conditions under which the Umegaki relative entropy and the transition probability $S_{1/2}$ is preserved under a quantum channel $\Phi$ for a pair of normal states $\sigma, \rho$ on a von Neumann algebra $\mathcal{M}$. It turned out that as in the classical case, this happens if and only if $\Phi$ is sufficient with respect to the experiment $(\mathcal{M}, \{\rho, \sigma\})$. Another equivalent condition is given in terms of the Connes cocycle derivative $[D\rho, D\sigma]_t$. In [66], the channels are not necessarily completely positive, only 2-positivity is assumed.

A quantum version of the factorization criterion was proved in [59] for finite dimensional algebras and in [42, 43] for all type I von Neumann algebras. In the paper [34], characterizations by preservation of other information quantities such as quantum $f$-divergences, Chernoff and Hoeffding distances are given. In particular, a characterization in terms of quantum $\alpha$-divergences for $\alpha \in (-3, 3)$ holds, see also [43]. Shirokov [74] studied sufficiency of bosonic channels.

The factorization criterion shows that sufficiency of a channel has strong implications on its structure and also on the structure of the involved states. For this reason, sufficiency is useful for finding equality conditions in inequalities involving entropic quantities. Most notably, it was used for characterization of Markov triples by equality in strong subadditivity of entropy, see [33, 42].

## 3.5 The results

In the first two works, we study characterizations of sufficient quantum channels by preservation of quantities related to hypothesis testing and quantum Fisher information. The rest is devoted to comparison of quantum experiments and a quantum randomization criterion.

### 3.5.1 Content of the papers [CE1–4]

**[CE1]** This paper focuses on the special case of channels given by restriction of the states to a subalgebra $\mathcal{A} \subseteq B(\mathcal{H})$ for a finite dimensional Hilbert space $\mathcal{H}$. If such a channel is sufficient, the subalgebra is called sufficient as well. For a pair of states $\{\rho_0, \rho_1\}$, we study subalgebras with the property that the optimal error probabilities for discrimination between the restrictions of $\rho_0$ and $\rho_1$ are the same as for discrimination of the original states. Such subalgebras are called 2-sufficient. A necessary condition for 2-sufficiency is found and it is proved that it is equivalent to sufficiency in the following special cases:

1. if $\rho_0^{it}\mathcal{A}\rho_0^{it} \subseteq \mathcal{A}$ for all $t \in \mathbb{R}$,
2. if $\mathcal{A}$ is commutative
3. if $\rho_0$ and $\rho_1$ commute.

Moreover, 2-sufficiency of $\mathcal{A}^{\otimes n}$ with respect to $\{\rho_0^{n\otimes}, \rho_1^{n\otimes}\}$ for all $n$ is equivalent to sufficiency of $\mathcal{A}$ with respect to $\{\rho_0, \rho_1\}$. This extends a classical result [77], which says that 2-sufficiency is equivalent to sufficiency for classical statistical experiments.

**[CE2]** We explore various reversibility (sufficiency) conditions for a 2-positive trace preserving map. We give an example of a non-quadratic operator convex function $f$ which is strictly convex, but preservation of the corresponding $f$-divergence does not imply sufficiency. This result shows a difference from the commutative case and complements the results of [34]. We prove characterizations in terms of the operator $d(\rho, \sigma) = \sigma^{-1/2}\rho\sigma^{-1/2}$, which is a quantum version of the Radon-Nikodym derivative. We also obtain a factorization criterion of the form

$$\rho_\theta = \Phi^*(S_\theta)\rho, \qquad \theta \in \Theta,$$

where $\Phi^*$ is the adjoint of $\Phi$, $S_\theta$ is a positive operator satisfying $\Phi^*(S_\theta^2) = (\Phi^*(S_\theta))^2$ and $\rho$ is a fixed density operator. We further show that preservation of the $L_1$-distance, which is equivalent to 2-sufficiency, characterizes sufficiency if the experiment is extended to contain all orbits under the modular group of some dominating element. We also show that sufficiency is characterized by preservation of the Chernoff and Hoeffding distances and of a large class of quantum Fisher informations.

**[CE3]** In this paper, we investigate the quantum versions of the preorder $\preceq_2$ and deficiency $\delta_2$ with respect to testing problems. Characterizations of the two notions are found and it is proved that for binary quantum experiments, the two preorders are equivalent if and only if the more informative experiment $\mathcal{E}$ is abelian. Moreover, $\mathcal{F} \preceq_2 \mathcal{E}$ implies existence of a completely positive map $\mathcal{E} \to \mathcal{F}$, but this map is not necessarily trace preserving.

**[CE4]** For two channels $\Phi$ and $\Psi$ with the same input space, we define deficiency of $\Psi$ with respect to $\Phi$ as the smallest distance between $\Phi$ and post-processings of $\Psi$, that is,

$$\delta(\Psi, \Phi) = \inf_\alpha \|\alpha \circ \Psi - \Phi\|_\diamond,$$

where the infimum is taken over the set of all channels between the corresponding output spaces. Using the results of Chapter 4 on the properties of channels and the diamond norm, we prove that $\delta(\Psi, \Phi) \leq \epsilon$ if and only if for any ensemble $\{p_i, \rho_i\}_i$ of states on the input space coupled with an ancilla, the optimal success probabilities of the output ensembles satisfy

$$P_{succ}(\{p_i, (\Phi \otimes id)(\rho_i)\}_i) \leq P_{succ}(\{p_i, (\Psi \otimes id)(\rho_i)\}_i) + \frac{\epsilon}{2} P_{succ}(\{p_i, \rho_i\}_i).$$

As a consequence, we obtain a randomization criterion for arbitrary quantum statistical experiments in terms of optimal success probabilities for certain ensembles. Over previously known results, this has the advantage that the success probabilities have a clear operational meaning.

### 3.5.2 Conclusions and open problems

In this section, we proved various characterizations of the possibility of approximation of a quantum statistical experiment by randomizations of another, along the lines of the classical theory of statistical isomorphisms and comparison of statistical experiments. As it often happens, it is possible to obtain similar results as in the classical case, but the conditions are more strict, and also the proofs are quite different. For example, while sufficient classical channels are characterized by preservation of the $f$-divergence for an arbitrary strictly convex function $f$, this is no longer true in the quantum case. Even if $f$ is operator convex, the support of the representing measure must be large enough. Further, while the factorization criterion looks similar to the classical one, the fact that the two factors have to commute has some strong consequences for the structure of the states. On the other hand, it is probably not so surprising that entanglement (or some other form of additional information) is needed for the quantum randomization criterion.

It is still not clear whether preservation of error probabilities in hypothesis testing, or 2-sufficiency, is equivalent to sufficiency also in the quantum case. While it is true if this condition is required for testing of $n$ i.i.d. copies of the states for all $n$, the one shot condition seems to be not enough. But quantum experiments invariant under the modular group of a dominating element behave similarly to classical experiments in this respect.

As it was recently observed, Umegaki relative entropy for density operators on a separable Hilbert space is monotone under positive maps, [60]. It is also an interesting question whether an equality condition similar to sufficiency for positive maps can be proved.

The deficiency $\delta(\Phi, \Psi)$ for some special cases of channels already appeared in quantum information theory, for example in the definition of approximately (anti)degradable channels [80]. Our results can be further used to obtain an operational definition, similarly as it was done for antidegradable channels in [11]. Another possible application is to $\epsilon$-private and $\epsilon$-correctable channels [47].

The suggested framework can be applied to more general situations, for example for comparison of more specific quantum protocols, such as quantum combs [16] and more general kinds of processings. More precisely, the processing can consist of a combination of pre- and post-processing, also allowing some correlations between input and output systems, either classical or quantum. This would be closer to the original definition by Shannon, [72]. In is also possible to treat different types of positive maps, some results in this direction can be found in the preprint [41]. Although our methods rely on finite dimensions, it seems plausible that the useful properties of the norms can be extended also for channels operating on infinite dimensional Hilbert spaces.

The notion of strong and weak convergence and local asymptotic normality, related to Le Cam distance for quantum experiments is also worth investigation, see the joint paper with M. Guta [28].

### 3.5.3 Further related works by the author

- A. Jenčová, D. Petz, Sufficiency in quantum statistical inference, *Commun. Math. Phys.* **263** (2006), 259-276
- A. Jenčová , D. Petz, Sufficiency in quantum statistical inference. A survey with examples, *IDAQP* **9** (2006), 331-351
- M. Guta, A. Jenčová, Local asymptotic normality in quantum statistics, *Commun. Math. Phys.* **276** (2007), 341-379
- A. Jenčová, M.B. Ruskai, A unified treatment of convexity of relative entropy and related trace functions, with conditions for equality, *Rev. Math. Phys.* **22** (2010), 1099-1121
- A. Jenčová, D. Petz and J. Pitrik, Markov triplets on CCR-algebras, *Acta Sci. Math. (Szeged)*, **76** (2010), 27-50
- A. Jenčová, The structure of strongly additive states and Markov triplets on the CAR algebra, *J. Math. Phys.* **51** (2010), 112103
- A. Jenčová, Preservation of a quantum Rényi relative entropy implies existence of a recovery map, *J. Phys. A: Math. Theor.*, **50** (2017), 085303
- A. Jenčová, Rényi relative entropies and noncommutative Lp-spaces, arxiv:1609.08462, 2016

# Chapter 4

# Generalized quantum channels and measurements

In this chapter, we study measurements on quantum channels and other quantum devices, using their convex structure. We will concentrate on quantum systems represented on finite dimensional Hilbert spaces, where the sets of devices can be identified with convex subsets of quantum state spaces. Moreover, we will discuss only measurements with a finite set of outcomes.

Let $\mathcal{H}$ and $\mathcal{K}$ be finite dimensional Hilbert spaces. Let $\mathcal{L}(\mathcal{H},\mathcal{K})$ denote the vector space of linear maps $B(\mathcal{H}) \to B(\mathcal{K})$ and let $\mathcal{C}(\mathcal{H},\mathcal{K})$ be the set of channels. For $\Phi \in \mathcal{L}(\mathcal{H},\mathcal{K})$, the *Choi representation* [17] is defined as

$$C(\Phi) = \frac{1}{d}\sum_{i,j} \Phi(|e_i\rangle\langle e_j|) \otimes |e_i\rangle\langle e_j|,$$

where $d = \dim(\mathcal{H})$ and $|e_1\rangle, \dots, |e_d\rangle$ is some fixed orthonormal basis of $\mathcal{H}$. The map $\Phi \mapsto C(\Phi)$ is a linear isomorphism of $\mathcal{L}(\mathcal{H},\mathcal{K})$ onto $B(\mathcal{K}\otimes\mathcal{H})$. Moreover, the map $\Phi$ is completely positive if and only if $C(\Phi)$ is positive and $\Phi$ preserves trace if and only if $\mathrm{Tr}_{\mathcal{K}} C(\Phi) = \dim(\mathcal{H})^{-1} I_{\mathcal{H}}$. It follows that the Choi representation identifies the set $\mathcal{C}(\mathcal{H},\mathcal{K})$ with a compact convex subset of the bipartite state space $\mathcal{S}(\mathcal{K}\otimes\mathcal{H})$.

## 4.1 Quantum channel measurements

A natural implementation of a channel measurement is obtained by applying the channels on an input state and measure the outcome by a POVM. A more general scheme can be described by a triple $(\mathcal{H}_0, \rho, M)$, where $\mathcal{H}_0$ is a (finite dimensional) ancilla, $\rho \in \mathcal{S}(\mathcal{H}\otimes\mathcal{H}_0)$ and $M$ is a POVM on $\mathcal{K}\otimes\mathcal{H}_0$. In fact, this is the most general form of a channel measurement considered in the literature.

Let $X$ be the (finite) set of outcomes. The outcome probabilities are given as

$$p_x(\Phi) = \mathrm{Tr}\,(\Phi\otimes id_{\mathcal{H}_0})(\rho)M_x, \qquad x \in X. \tag{4.1}$$

Such measurements serve as decision rules in statistical decision problems for quantum channels. For example, the discrimination problem of Example 4 can be formulated for a pair of channels $\Phi_0, \Phi_1$ in an obvious way. Decision rules for this problem are given by triples $(\mathcal{H}_0, \rho, M)$ with two-outcome POVMs $\{M, I - M\}$. Let $\lambda$, $1-\lambda$ be prior probabilities, then the Bayes error probability is given by

$$\begin{aligned} P_e(\mathcal{H}_0, \rho, M) &:= \lambda \mathrm{Tr}\,(\Phi_0 \otimes id_{\mathcal{H}_0})(\rho)M + (1-\lambda)\mathrm{Tr}\,(\Phi_1 \otimes id_{\mathcal{H}_0})(\rho)(I - M) \\ &= (1-\lambda) + \mathrm{Tr}\,((\lambda\Phi_0 - (1-\lambda)\Phi_1) \otimes id)(\rho)M. \end{aligned}$$

Optimizing over all $(\mathcal{H}_0, \rho, M)$, we obtain

$$\min P_e(\mathcal{H}_0, \rho, M) = \frac{1}{2}(1 - \|\lambda\Phi_0 - (1-\lambda)\Phi_1\|_\diamond),$$

where $\|\cdot\|_\diamond$ is the diamond norm, given by [45]

$$\|\Phi\|_\diamond := \sup_{\rho \in \mathcal{S}(\mathcal{H}\otimes\mathcal{H})} \|(\Phi \otimes id)(\rho)\|_1 \tag{4.2}$$

for any $\Phi \in \mathcal{L}(\mathcal{H}, \mathcal{K})$. For more information on discrimination of channels and the diamond norm, see [83].

The description of measurement by triples is not unique. It is easy to see that there are many different triples with the same outcome probabilities. A different framework was introduced in [84], where the *process POVMs* were introduced. Similarly to a POVM, a process POVM (with a finite set of outcomes) is defined as a collection of positive operators on $\mathcal{K} \otimes \mathcal{H}$, summing up to $I_\mathcal{K} \otimes \sigma$, where $\sigma$ is some element in $\mathcal{S}(\mathcal{H})$. The outcome probabilities are given by

$$p_x(\Phi) = \mathrm{Tr}\, C(\Phi)(\rho)F_x, \qquad x \in X.$$

For each triple $(\mathcal{H}_0, \rho, M)$, there is a unique process POVM with the same outcome probabilities for all channels. Conversely, for each process POVM, there are many corresponding triples. This description of measurements is again not one-to-one: one can see that there exist different process POVMs describing the same channel measurement.

## 4.2 Quantum combs and testers

Process POVMs belong to a more general formalism of (probabilistic) *quantum combs*, see [16, 15] for details. Quantum combs are used in description of quantum networks. This formalism is hierarchical and contains also admissible transformations between networks, these correspond to networks of a higher rank.

A formal definition of a quantum comb is recursive: a (deterministic) 1-comb is a Choi matrix of a channel, an $N$-comb is the Choi matrix of a completely positive map on a multipartite tensor product that maps $(N-1)$-combs to 1-combs. A probabilistic $N$-comb is a positive operator majorized by a deterministic $N$-comb. All quantum combs can be represented by memory

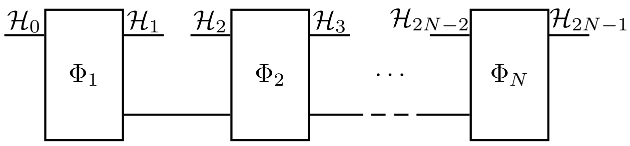


Figure 4.1: A deterministic quantum $N$-comb

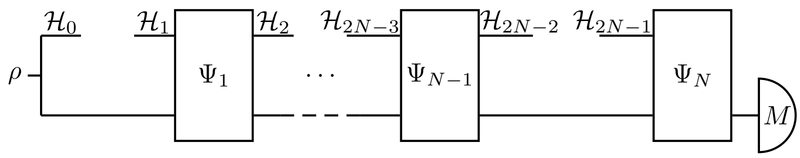


Figure 4.2: Quantum $N$-tester

channels, which are given by a sequence of channels connected by an ancilla, these form the "teeth" of the comb, see Fig. 4.1.

*Quantum testers* are a special case of probabilistic quantum combs. More precisely, a quantum $N$-tester $T = \{T_1, \ldots, T_m\}$ is a collection of positive operators summing up to a special type of $(N+1)$-comb which maps all $N$-combs to 1. $N$-testers describe measurements on deterministic $N$-combs and can be represented as a deterministic comb with a measurement on the ancilla, see Fig. 4.2. Similarly as for quantum channels, the optimal Bayes error probabilities for discrimination of quantum $N$-combs $\Xi_0$ and $\Xi_1$ with prior probabilities $\lambda, 1-\lambda$ have the form

$$\min_T P_e(T) = \frac{1}{2}(1 - \|\lambda\Xi_0 - (1-\lambda)\Xi_1\|_{\diamond N}),$$

where

$$\|\Xi\|_{\diamond N} := \sup_T \|(T_0+T_1)^{1/2}\Xi(T_0+T_1)^{1/2}\|_1$$

for any (Hermitian) matrix $\Xi$. The supremum is taken over the set of all two-outcome $N$-testers $T = \{T_0, T_1\}$. See also [30, 29] for a similar framework of quantum games.

## 4.3 The results

The quantum devices described above have a natural convex structure which reflects the possibility to switch randomly between different devices of the same type. As it is in the case of states, a measurement on any convex set $K$ can be defined as an affine map from $K$ into some classical state space, assigning to each element the corresponding outcome probabilities. In the works below, we study the relation between this definition of measurements or, more generally, affine maps into state spaces, and the framework of quantum combs and testers. We also show that the distinguishability norms $\|\cdot\|_{\diamond N}$ arise naturally from the structure of quantum channels and combs as convex subsets of multipartite state spaces.

### 4.3.1 Content of the papers [CS1], [CS2]

**[CS1]** In this paper, we study affine maps on a convex subset $K$ of the state space $\mathcal{S}(\mathcal{A})$ of a finite dimensional C*-algebra $\mathcal{A}$ into the state space of another finite dimensional C*-algebra $\mathcal{B}$. If $\mathcal{B}$ is commutative, such a map corresponds to a measurement. We prove that all measurements extend to positive affine maps on $\mathcal{S}(\mathcal{A})$ if and only if $K$ is a section of the state space, that is

$$K = K^{\perp\perp} \cap \mathcal{S}(\mathcal{A}).$$

In this case, the measurements are defined by collections of positive operators satisfying certain normalization condition. Such collections are called *generalized POVMs* with respect to $K$. This is true for the sets of $N$-combs, since these sets are given by linear constraints. The corresponding generalized POVMs for the set of $N$-combs are exactly the $N$-testers, in particular, for quantum channels we obtain the process POVMs. The relation between measurements and generalized POVMs is not one-to-one, as different generalized POVMs may correspond to the same measurement.
If $\mathcal{B}$ is not commutative, we also require that the affine map extends to a completely positive map on the subspace generated by $K$. We show that each such map is the restriction of a completely positive map on $\mathcal{A}$, called a generalized channel. If the set $K$ contains the tracial state, the set of generalized channels forms a section of a multipartite state space. This leads to a definition of a generalized supermap, special cases of which are quantum combs and testers. We also discuss the equivalence relation on generalized channels and POVMs, given by restriction to $K$.

**[CS2]** We continue the study of sections of quantum state spaces or more general bases of the cone of positive operators on a (finite dimensional) Hilbert space. It is shown that the section defines a norm in the space of Hermitian operators, which is a distinguishability norm for elements of the section. The dual norm is studied and it is shown that it again corresponds to a base section. These norms are a generalization of both the base norms and the order unit norms and have similar properties. It is proved that for the set of channels, the corresponding norm is the diamond norm and (logarithm of) the dual norm is the conditional max-relative entropy. Similarly, for $N$-combs, we obtain the norm $\|\cdot\|_{\diamond N}$ and the dual norm also has a similar form.
We further study statistical decision problems for elements of the section. It is shown that average risks (or payoffs) of decision rules can be expressed in terms of the norm corresponding to a related base section, in particular, the dual of the diamond norm can be used to express optimal risks for decision rules for the set of quantum states. Optimality conditions for decision rules are also given. As a corollary, a necessary and sufficient condition is obtained, under which there is a triple with a maximally entangled input state which is optimal for discrimination of two channels.

### 4.3.2 Conclusions and open problems

We investigated affine maps on convex subsets of state spaces and proved that such maps represent certain quantum devices. Moreover, it was shown that some important norms on the set of Hermitian linear maps can be obtained from the convex structure of sets of quantum channels and that optimal risks of decision rules can be given in terms of similar norms. In particular, we proved that the dual of the diamond norm gives the maximum success probability in multiple hypothesis testing problems, this was crucial for the proof of the quantum randomization criterion in the previous chapter.

In the recent paper [44], optimality conditions for multiple hypothesis testing problems on quantum channels were re-obtained, using semidefinite programming. Conditions for existence of an optimal test with maximally entangled input states were again discussed and an upper bound on error probability was given in the case that these conditions are not satisfied. One can see that similar conditions and bounds can be found for more general quantum combs. A specific example of a comb is an i.i.d. sequence of channels. It is a question whether the mentioned results can be used to investigate asymptotic properties of channels.

### 4.3.3 Further related works by the author

- A. Jenčová, Extremality conditions for generalized channels, *J. Math. Phys.* **53** (2012), 122203
- A. Jenčová, Extremal generalized quantum measurements, *Linear Algebra Appl.* **439** (2013), 4070-4079
- Z. Puchala, A. Jenčová, M. Sedlák, M. Ziman, Exploring boundaries of quantum convex structures: special role of unitary processes, *Phys. Rev. A* **92** (2015), 012304
- A. Jenčová, On the convex structure of process POVMs, *J. Math. Phys.* **57** (2016), 015207
- A. Jenčová, M. Plávala, Conditions for optimal input states for discrimination of quantum channels, *J. Math. Phys.* **57** (2016), 122203

# Bibliography


[1] S. I. Amari, *Differential-geometrical Methods in Statistics*, Lecture Notes in Statistics, Springer-Verlag, New York (1985)

[2] S. I. Amari, $\alpha$-divergence is unique, belonging to both $f$-divergence and Bregman divergence classes, *IEEE Trans. Inf. Theor.*, **55** (2009), 2009

[3] S. I. Amari, H. Nagaoka, *Method of Information Geometry*, AMS Monograph, Oxford University Press (2000)

[4] H. Araki, Relative entropy of states of von Neumann algebras, *Publ. RIMS Kyoto Univ.*, **11** (1976), 809–833

[5] V. P. Belavkin, P. C. Staszewski, C$*$-algebraic generalizations of relative entropy and entropy, *Ann. Inst. Henri Poincaré A*, **37** (1982), 51–58

[6] D. Blackwell, Comparison of experiments, *Proc. 2nd Berkeley Symp. on Math. Stat. and Probab.*, (1951), 93–102

[7] F. Buscemi, Comparison of Quantum Statistical Models: Equivalent Conditions for Sufficiency, *Commun. Math. Phys.*, **310** (2012), 625–647

[8] F. Buscemi, Fully quantum second-law–like statements from the theory of statistical comparisons, arXiv:1505.00535, (2014)

[9] F. Buscemi, Degradable channels, less noisy channels, and quantum statistical morphisms: an equivalence relation, *Problems of Information Transmission*, **53** (2016), 201-213

[10] F. Buscemi, N. Datta, Equivalence between divisibility and monotonic decrease of information in classical and quantum stochastic processes, *Phys. Rev. A*, **93** (2016), 012101

[11] F. Buscemi, N. Datta, S. Strelchuk, Game-theoretic characterization of antidegradable channels, *J. Math. Phys.*, **55** (2014), 092202

[12] N. N. Cencov, *Statisticheskie reshayushhie pravila i optimal'nye vyvody*, Nauka, Moscow (1972), Engl. transl.: 1982, *Statistical decision rules and optimal inference*, American Mathematical Society, Providence

[13] N. N. Cencov, E. A. Morozova, Markov invariant geometry on state manifolds, *Itogi Nauki i Tekhniki*, **36** (1990), 69–102

[14] A. Chefles, The Quantum Blackwell Theorem and Minimum Error State Discrimination, arXiv:0907.0866, (2009)

[15] G. Chiribella, G. M. D'Ariano, P. Perinotti, Memory effects in quantum channel discrimination, *Phys. Rev. Lett.*, **101** (2008)

[16] G. Chiribella, G. M. D'Ariano, P. Perinotti, Theoretical framework for quantum networks, *Phys. Rev. A*, **80** (2009)

[17] M. D. Choi, Completely Positive Linear Maps on Complex matrices, *Lin. Alg. Appl.*, **10** (1975), 285–290

[18] I. Csiszár, Information type measure of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.*, **2** (1967), 299–318

[19] I. Csiszár, J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Channels*, Academic Press, New York - San Francisco - London (1981)

[20] A. P. Dawid, On the concepts of sufficiency and ancillarity in the presence of nuisance parameters, *J. R. Stat. Soc. B*, **37** (1975), 248–258

[21] B. Efron, Defining the curvature of a statistical problem (with applications to second order efficiency), *Annals of Statistics*, **3** (1975), 1189–1242

[22] S. Eguchi, Second order efficiency of minimum contrast estimators in a curved exponential family, *Ann. Statist.*, **11** (1983), 793

[23] P. Gibilisco, T. Isola, Connections on statistical manifolds of density operators by geometry of non-commutative $L_p$-spaces, *IDAQP*, **2** (1999), 169–178

[24] P. Gibilisco, G. Pistone, Connections on non-parametric statistical manifolds by Orlicz space geometry, *IDAQP*, **1** (1998), 325–347

[25] M. R. Grasselli, Dual connections in nonparametric classical information geometry, *Ann. Inst. Stat. Math.*, **62** (2010), 873–896

[26] M. Graselli, R. Streater, The quantum information manifold for $\epsilon$-bounded forms, *Rep. Math. Phys.*, **46** (2000), 325–335

[27] M. Grasselli, R. Streater, On the uniqueness of the Chentsov metric in quantum information geometry, *IDAQP*, **4** (2001), 173–182,

[28] M. Guta, A. Jenčová, Local asymptotic normality in quantum statistics, *Commun. Math. Phys.*, **276** (2007), 341–379

[29] G. Gutoski, On a measure of distance for quantum strategies, *J. Math. Phys.*, **53** (2012)

[30] G. Gutoski, J. Watrous, Toward a General Theory of Quantum Games, In: *STOC 07: Proceedings of the 39th annual ACM Symposium on Theory of Computing*, Annual ACM Symposium on Theory of Computing (2007), 565–574

[31] H. Hasegawa, Dual geometry of the Wigner–Yanase–Dyson information content, *IDAQP*, **6** (2003), 413

[32] H. Hasegawa, D. Petz, D. Non-commutative extension of information geometry II, In: O. H. et. al. (ed.), *Quantum Communication, Computing and Measurement*, Plenum Press, New York (1997)

[33] P. Hayden, R. Jozsa, D. Petz, A. Winter, Structure of states which satisfy strong subadditivity of quantum entropy with equality, *Commun. Math. Phys.*, **246** (2004), 359–374

[34] F. Hiai, M. Mosonyi, D. Petz, C. Beny, Quantum f-divergences and error correction, *Rev. Math. Phys.*, **23** (2011), 691–747,

[35] C. E. Helstrom, *Quantum Detection and Estimation Theory* , Academic Press, New York (1976)

[36] A. S. Holevo, Statistical decision theory for quantum systems, *Journal of Multivariate Analysis*, **3** (1973), 337 – 394

[37] A. S. Holevo, *Statistical Structure of Quantum Theory* , Springer-Verlag, Berlin Heidelberg (2001)

[38] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory (2-nd English ed.)*, Edizioni della Normale, Pisa (2011)

[39] H. Jeffreys, An invariant form for the prior probability in estimation problems, *Proceedings of Royal Society A*, **186** (1946), 453–461

[40] A. Jenčová, Comparison of quantum binary experiments, *Rep. Math. Phys.*, **70** (2012), 237–249

[41] A. Jenčová, Randomization theorems for quantum channels, arXiv:1404.3900 (2014)

[42] A. Jenčová, D. Petz, Sufficiency in quantum statistical inference, *Commun. Math. Phys.*, **263** (2006), 259–276,

[43] A. Jenčová, D. Petz, Sufficiency in quantum statistical inference. A survey with examples, *Inf. Dim. Anal. Quant. Prob. Rel. Top.*, **9** (2006), 331–351

[44] A. Jenčová, M. Plávala, Conditions for optimal input states for discrimination of quantum channels,, *J. Math. Phys.*, **57**, 122203 (2016)

[45] A. Kitaev, Quantum computations: Algorithms and error correction, *Russian Mathematical Surveys*, **52** (1997), 1191–1249

[46] J. Körner, K. Marton, Comparison of two noisy channels, In: *Colloquia Mathematica Societatis Janos Bolyai, Topics in Information Theory*, North Holland (1977), 411–424

[47] D. Kretschmann, D. W. Kribs, R. W. Spekkens, Complementarity of private and correctable subsystems in quantum cryptography and error correction, *Phys. Rev. A*, **78** (2008), 032330,

[48] S. Kullback, R. A. Leibler, On information and sufficiency, *Ann. Math. Stat.*, **22** (1951), 78–86

[49] L. E. Labuschagne, A crossed product approach to Orlicz spaces, *Proc. London Math. Soc.*, **107** (2013), 965–1003

[50] S. L. Lauritzen, Statistical manifolds, In: S. Gupta (ed.), *Differential geometry in statistical inference*, Institute of Mathematical Statistics, Hayward (1987)

[51] L. Le Cam, Sufficiency and approximate sufficiency, *Ann. Math. Statist.*, **34** (1964), 1419–1455

[52] L. Le Cam, *Asymptotic Methods in Statistical Decision Theory*, Springer Verlag, New York (1986)

[53] A. Lesniewski, M. Ruskai, Monotone Riemannian metrics and relative entropy on noncommutative probability spaces, *J. Math. Phys.*, **40** (1999), 5702–5724

[54] F. Liese, I. Vajda, *Convex Statistical Distances*, Teubner-Texte zur Mathematik, Leipzig (1987)

[55] W. A. Majewski, L. Labuschagne, On applications of Orlicz spaces to Statistical Physics, *Ann. H. Poincare*, **15** (2014), 1197–1221

[56] T. Masuda, $L_p$-spaces for von Neumann algebra with reference to a faithful normal semifinite weight, *Publ. RIMS, Kyoto Univ.*, **19** (1983), 673–727

[57] K. Matsumoto, A quantum version of randomization condition, arXiv:1012.2650 (2010)

[58] K. Matsumoto, An example of a quantum statistical model which cannot be mapped to a less informative one by any trace preserving positive map, arXiv:1409.5658 (2014)

[59] M. Mosonyi, D. Petz, Structure of sufficient quantum coarse-grainings, *Lett. Math. Phys.*, **68** (2004), 19–30,

[60] A. Müller-Hermes, D. Reeb, Monotonicity of the quantum relative entropy under positive maps, arXiv:1512.06117 (2015)

[61] H. Nagaoka, Differential geometrical aspects of quantum state estimation end relative entropy, In: R. L. H. V. P. Belavkin, O. Hirota (ed.), *Quantum Communication, Computing and Measurement*, Plenum Press, New York (1995)

[62] H. Nagaoka, S. I. Amari, Differential geometry of smooth families of probability distributions, Tech. Rep. METR 82-7, University of Tokyo, Tokyo (1982)

[63] D. Petz, Quasi-entropies for finite quantum systems, *Rep. Math. Phys.*, **23** (1984), 57–65

[64] D. Petz, Quasi-entropies for states of a von Neumann algebra, *Publ. RIMS, Kyoto Univ.*, **21** (1985), 787–800

[65] D. Petz, Sufficient subalgebras and the relative entropy of states of a von Neumann algebra, *Commun. Math. Phys.*, **105** (1986), 123–131

[66] D. Petz, Sufficiency of channels over von Neumann algebras, *Quart. J. Math. Oxford*, **39** (1988), 97–108

[67] D. Petz, Monotone metrics on matrix spaces, *Lin. Alg. Appl.*, **244** (1996), 81–96

[68] G. Pistone, M. P. Rogantin, The exponential statistical manifold: mean parameters, orthogonality and space transformations, *Bernoulli*, **5** (1999), 721–760

[69] G. Pistone, C. Sempi, An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one, *Ann. Statist.*, **23** (1995), 1543–1561

[70] M. Raginsky, Shannon meets Blackwell and Le Cam: Channels, codes, and statistical experiments, *IEEE International Symposium on Information Theory Proceedings*, (2011), 1220–1224

[71] C. R. Rao, Information and accuracy attainable in the estimation of statistical parameters, *Bulletin of the Calcuta Mathematical Society*, **37** (1945), 81–91

[72] C. E. Shannon, A note on a partial ordering for communication channels, *Inform. Control*, **1** (1958), 390–397

[73] S. Sherman, On a theorem of Hardy, Littlewood, Pólya and Blackwell, *Proc. Nat. Acad. Sciences*, **37** (1951), 826–831

[74] M. E. Shirokov, Reversibility of a quantum channel: general conditions and their applications to Bosonic linear channels, *J. Math. Phys.*, **54** (2013), 112201

[75] E. Shmaya, Comparison of information structures and completely positive maps, *J. Phys. A-Math. Gen.*, **38** (2005), 9717–9727

[76] C. Stein, Notes on a Seminar on Theoretical Statistics. I. Comparison of experiments, Tech. Rep., University of Chicago (1951)

[77] H. Strasser, *Mathematical Theory of Statistics*, de Gruyter, Berlin, New York (1985)

[78] R. Streater, Quantum Orlicz spaces in information geometry, *Open Sys. Inf. Dyn.*, **11** (2004), 350

[79] R. F. Streater, The Information Manifold for Relatively Bounded Potentials, *Tr. Mat. Inst. Steklova*, **228** (2000), 217–235

[80] D. Sutter, V. B. Scholz, R. Renner, Approximate degradable quantum channels, *IEEE International Symposium on Information Theory Proceedings*, (2015), 2767–2771

[81] E. Torgersen, Comparison of statistical experiments when the parameter space is finite, *Z. Wahrscheintlichkeitstheorie verw. geb.*, **16** (1970), 219–249

[82] H. Umegaki, Conditional expectation in an operator algebra, IV: Entropy and information, *Kodai Math. Sem. Rep.*, **14** (1962), 59–85

[83] J. Watrous, Theory of Quantum Information, University of Waterloo, Waterloo (2011), http://www.cs.uwaterloo.ca/~watrous/CS766/

[84] M. Ziman, Process positive-operator-valued measure: A mathematical framework for the description of process tomography experiments, *Phys. Rev. A*, **77** (2008)

# Part II

# The papers

# Chapter 5

# The structures of information geometry


[IG1] A. Jenčová, Geometry of quantum states: dual connections and divergence functions, *Rep. Math. Phys.* **47** (2001), 121-138

[IG2] A. Jenčová, Generalized relative entropies as contrast functionals on density matrices, *Int. J. Theor. Phys.* **43**, (2004), 1635-1649

[IG3] A. Jenčová, Flat connections and Wigner-Yanase-Dyson metrics, *Rep. Math. Phys.* **52** (2003), 331–351

[IG4] A. Jenčová, Quantum information geometry and non-commutative $L_p$ spaces, *IDAQP* **8** (2005), 215–233

[IG5] A. Jenčová, A construction of a nonparametric quantum information manifold, *J. Funct. Anal.* **239** (2006), 1-20

[IG6] A. Jenčová, On quantum information manifolds, In: *Algebraic and Geometric Methods in Statistics*, Cambridge University Press 2010

# GEOMETRY OF QUANTUM STATES: DUAL CONNECTIONS AND DIVERGENCE FUNCTIONS

ANNA JENČOVÁ

Mathematical Institute, Slovak Academy of Sciences,
Štefánikova 49, SK-814 73 Bratislava, Slovakia
(e-mail: JENCA@mat.savba.sk)

*(Received February 21, 2000 — Revised May 23, 2000)*

In a finite quantum state space with a monotone metric, a family of torsion-free affine connections is introduced, in analogy with the classical $\alpha$-connections defined by Amari. The dual connections with respect to the metric are found and it is shown that they are, in general, not torsion-free. The torsion and the Riemannian curvature are computed and the existence of efficient estimators is treated. Finally, geodesics are used to define a divergence function.

## 1. The classical case

Let $\mathcal{S} = \{p(x,\theta) \mid \theta \in \Theta \subseteq \mathbb{R}^m\}$ be a smooth family of classical probability distributions on a sample space $\mathcal{X}$. Then $\mathcal{S}$ can naturally be viewed as a differentiable manifold. The differential-geometrical aspects of a statistical manifold and their statistical implications were studied by many authors. The Riemannian structure is given by the Fisher information metric tensor

$$g_{ij}(\theta) = E_\theta[\partial_i \log p(x,\theta)\partial_j \log p(x,\theta)],$$

where $\partial_i$ denotes $\frac{\partial}{\partial\theta_i}$. In 1972, Chentsov in [6] introduced a family of affine connections in $\mathcal{S}$ and proved that the Fisher information and these connections were unique (up to a constant factor) in the manifold of distributions on a finite number of atoms, in the sense that these are invariant with respect to transformations of the sample space. In [1], Amari defined a one-parameter family of $\alpha$-connections in $\mathcal{S}$, which turned out to be the same as those defined by Chentsov. They may be introduced using the following $\alpha$-representations of the tangent space:

Let $g_\alpha$ be a one-parameter family of functions, given by

$$g_\alpha(x) = \begin{cases} \dfrac{2}{1-\alpha} x^{\frac{1-\alpha}{2}} & \text{for } \alpha \neq 1 \\ \log x & \text{for } \alpha = 1. \end{cases}$$

Let $l_\alpha(x,\theta) = g_\alpha(p(x,\theta))$. The vector space spanned by the functions $\partial_i l_\alpha(x,\theta)$, $i = 1,\dots,p$, is called the $\alpha$-representation of the tangent space. The metric tensor $g_{ij}$ is then

$$g_{ij}(\theta) = \int \partial_i l_\alpha(x,\theta)\partial_j l_{-\alpha}(x,\theta)dP.$$

The $\alpha$-connections are defined by

$$\Gamma^\alpha_{ijk}(\theta) = \int \partial_i\partial_j l_\alpha(x,\theta)\partial_k l_{-\alpha}(x,\theta)dP.$$

From this, it is clear that these connections are torsion-free, i.e. $S^\alpha_{ijk} = \Gamma^\alpha_{ijk} - \Gamma^\alpha_{jik} = 0$, $\forall i, j, k$, $\forall\theta$. Let now $\nabla$ and $\nabla^*$ be two covariant derivatives on $\mathcal{S}$, then we say that the covariant derivatives (the affine connections) are dual with respect to the metric if

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z)$$

or, in coefficients,

$$\partial_i g_{jk} = \Gamma_{ijk} + \Gamma^*_{ikj}.$$

It is easy to see that the $\alpha$ and $-\alpha$ connections are mutually dual.

Further, let $\eta$ be another coordinate system in $\mathcal{S}$. The natural basis of the tangent space $T_P$ at $P \in \mathcal{S}$ is $\{\partial_i\}$, $\partial_i = \frac{\partial}{\partial\theta_i}$ for the coordinate system $\theta$ and $\{\partial^i\}$, $\partial^i = \frac{\partial}{\partial\eta_i}$ for $\eta$. We say that $\theta$ and $\eta$ are mutually dual if their natural bases are biorthogonal, i.e. if

$$g(\partial_i, \partial^j) = \delta_i^j.$$

The metric tensor in the basis $\{\partial^i\}$ is given by

$$g(\partial^i, \partial^j) = g^{ij}, \qquad (g^{ij}) = (g_{ij})^{-1}.$$

The next theorem (Theorem 3.4 in [1]) gives the necessary and sufficient condition for existence of such pair of coordinate systems.

THEOREM 1.1. *If a Riemannian manifold $\mathcal{S}$ has a pair of dual coordinate systems $(\theta,\eta)$, then there exist potential functions $\psi(\theta)$ and $\phi(\eta)$ such that*

$$g_{ij} = \partial_i\partial_j\psi(\theta), \qquad g^{ij} = \partial^i\partial^j\phi(\eta). \tag{1}$$

*Conversely, if either of the potential functions $\psi$ or $\phi$ exist such that* (1) *holds, there exists a pair of dual coordinate systems. The dual coordinate systems are related by the Legendre transformations*

$$\theta_i = \partial^i\phi(\eta), \qquad \eta_i = \partial_i\psi(\theta)$$

*and the two potential functions satisfy the identity*

$$\psi(\theta) + \phi(\eta) - \sum_i \theta_i\eta_i = 0. \tag{2}$$

The most interesting results of [1] concern the case when the manifold $\mathcal{S}$ is $\alpha$-flat, i.e. the Riemannian curvature tensor of the $\alpha$-connection vanishes. Then $\mathcal{S}$ is also $-\alpha$-flat. It is known that for flat manifolds, an affine coordinate system exists, i.e. the coefficients of the connection vanish. The next theorem (Theorem 3.5 in [1]) reveals the dualistic structure of $\alpha$-flat manifolds.

THEOREM 1.2. *When a Riemannian manifold $S$ is flat with respect to a pair of torsion-free dual affine connections $\nabla$ and $\nabla^*$, there exists a pair $(\theta, \eta)$ of dual coordinate systems such that $\theta$ is a $\nabla$-affine and $\eta$ is a $\nabla^*$-affine coordinate system.*

This result can be directly applied to the exponential and mixture families

$$p(x,\theta) = \exp\left\{\sum_i \theta_i c_i(x) - \psi(\theta)\right\}$$

and

$$p(x,\theta) = \sum_i \theta_i c_i(x) + \left(1 - \sum_i \theta_i\right) c_{n+1}(x)$$

which are $\pm 1$-flat, and the extended $\alpha$-families

$$l_\alpha(x,\theta) = \sum_i^{n+1} \theta_i c_i(x)$$

which are $\pm\alpha$-flat (note that the extended $\alpha$-families are not normed to 1).

Let us consider an $\alpha$-flat family $\mathcal{S}$ with the dual coordinate systems $(\theta, \eta)$ and let $\psi(\theta)$ and $\phi(\eta)$ be the potential functions. In [1], a divergence is introduced in $\mathcal{S}$. It is called the $\alpha$-divergence and it is given by

$$D_\alpha(\theta, \theta') = \psi(\theta) + \phi(\eta') - \sum_i \theta_i \eta'_i.$$

The divergence is not a usual distance, but it has some important properties:
(i) it is strictly positive, $D_\alpha(\theta, \theta') \geq 0$ with $D_\alpha(\theta, \theta') = 0$ iff $\theta = \theta'$,
(ii) it is jointly convex in $\theta$ and $\theta'$,
(iii) $D_\alpha(\theta, \theta') = D_{-\alpha}(\theta', \theta)$,
(iv) it satisfies the relation

$$D_\alpha(\theta, \theta + d\theta) = D_\alpha(\theta + d\theta, \theta) = \frac{1}{2}\sum_{ij} g_{ij}(\theta) d\theta_i d\theta_j,$$

hence it induces the Riemannian distance, given by the Fisher information.

Moreover, the divergences are shown to satisfy a generalized Pythagorean relation.

There were some attempts to introduce a similar family of affine structures in the differentiable manifold of states of an $n$-level quantum system with a monotone metric. The main difficulty here is that, as was first shown in [7], there is no unique quantum analogue of the Fisher information metric and, moreover, the connections are in general not torsion-free. Hasegawa in [9] studied the case of quantum exponential and mixture families with the Kubo–Mori metric. In [13], the exponential and mixture connections (i.e. the case $\alpha = \pm 1$) were defined also for arbitrary monotone metric and a divergence function was introduced. The aim of this paper is to use a similar method to define the $\alpha$-connections and divergence functions for each $\alpha$ and to investigate the dualistic properties of the manifold.

## 2. The state space

Let $\mathcal{M}$ denote the differentiable manifold of all $n$-dimensional complex hermitian matrices and let $\mathcal{M}^+ = \{M \in \mathcal{M} \mid M > 0\}$. Let $\tilde{T}_M$ be the tangent space at $M$, then $\tilde{T}_M$ can be identified with $\mathcal{M}$ considered as a vector space. We introduce a Riemannian structure in $\mathcal{M}^+$, defining an inner product in $\tilde{T}_M$ by

$$\lambda_M(X, Y) = \operatorname{Tr} X J_M(Y),$$

where $J_M$ is a suitable superoperator on matrices. The state space of an $n$-level quantum system can be identified with the submanifold

$$\mathcal{D} = \{D \in \mathcal{M}^+ \mid \operatorname{Tr} D = 1\}.$$

The tangent space $T_D \subset \tilde{T}_D$ is the real vector space of all self-adjoint traceless matrices. If we consider the restriction of the Riemannian structure $\lambda$ onto $\mathcal{D}$, it is natural to require that $\lambda$ be monotone, in the sense that if $T$ is a stochastic map, then

$$\lambda_{T(M)}(T(X), T(X)) \leq \lambda_M(X, X), \qquad \forall M \in \mathcal{M}^+, \ X \in \mathcal{M}.$$

As it was proved in [14], this is true iff $J_M$ is of the form

$$J_M = (R_M^{\frac{1}{2}} f(L_M R_M^{-1}) R_M^{\frac{1}{2}})^{-1}, \tag{3}$$

where $f : R^+ \to R$ is an operator monotone function such that $f(t) = t f(t^{-1})$ for every $t > 0$ and $L_M(A) = MA$ and $R_M(A) = AM$, for each matrix $A$. We also adopt a normalization condition $f(1) = 1$. Notice that we have $J_M(X) = M^{-1}X$ whenever $X$ and $M$ commute, in particular, $J_M(M) = I$.

Let $T_D^*$ be the cotangent space of $\mathcal{D}$ at $D$, then $T_D^*$ is the vector space of all observables $A$ with the zero mean at $D$ (i.e. $\operatorname{Tr} DA = 0$). It is easy to see that

$$T_D^* = \{J_D(H) \mid H \in T_D\}. \tag{4}$$

The metric $\lambda$ induces an inner product in $T_D^*$, namely,

$$\varphi(A_1, A_2) = \lambda(J_D^{-1}(A_1), J_D^{-1}(A_2)) = \operatorname{Tr} A_1 J_D^{-1}(A_2).$$

It can be interpreted as a generalized covariance of the observables $A_1$ and $A_2$.

EXAMPLE 2.1. Let the metric be determined by $J_D(H) = G$, where $GD+DG = 2H$, then it is called the metric of the symmetric logarithmic derivative. This metric is monotone, with the corresponding operator monotone function $f(x) = \frac{1+x}{2}$, see [5, 11, 15].

EXAMPLE 2.2. Another important example of a monotone metric is the well-known Kubo–Mori metric determined by $J_D(H) = \frac{d}{dt}\log(D+tH)|_{t=0}$. The monotone metric is given by

$$\lambda_D(H, K) = \operatorname{Tr} H J_D(K) = \frac{\partial^2}{\partial t \partial s} \operatorname{Tr}(D + tH)\log(D + sK)|_{t,s=0}.$$

Note that this metric is induced by the relative entropy $D(\rho, \sigma) = \operatorname{Tr}\rho(\log\rho - \log\sigma)$ when the density matrices $\rho$ and $\sigma$ are infinitesimally distant from each other.

EXAMPLE 2.3. [11, 16] Let $J_D(H) = \frac{1}{2}(D^{-1}H + HD^{-1})$. The corresponding metric is monotone, with $f(x) = \frac{2x}{x+1}$, and it is called the metric of the right logarithmic derivative.

For more about monotone metrics and their use see [15, 16, 10].

## 3. The $g$-representation

Let $g : \mathbb{R} \to \mathbb{R}$ be a smooth (strictly) monotone function. We define an operator $L_g[M] : \tilde{T}_M \to \tilde{T}_{g(M)}$ by

$$L_g[M](H) = \frac{d}{ds} g(M + sH)|_{s=0}.$$

The following Lemma was proved in [17]. As it is frequently used in the sequel, we repeat the proof here.

LEMMA 3.1.

(i) *$L_g[M]$ is a linear map.*

(ii) *$L_{f\circ g}[M] = L_f[g(M)]L_g[M]$. In particular, if $g$ is invertible then $L_g[M]$ is invertible and $L_g[M]^{-1} = L_{g^{-1}}[g(M)]$.*

(iii) *$L_g[M]$ is self-adjoint, with respect to the inner product $\langle A, B\rangle = \operatorname{Tr} AB$ in $\mathcal{M}$.*

(iv) *$L_g[M](M) = g'(M)M$, where $g'(x) = \frac{d}{dx}g(x)$.*

(v) *$L_g[M]^{-1}(I) = (g'(M))^{-1}$, $I$ is the identity matrix.*

*Proof:* It is convenient to use the orthogonal (with respect to the inner product $\langle A, B\rangle$ above) decomposition of the tangent space introduced in [9], $\tilde{T}_M = C(M) \oplus C(M)^{\perp}$, where $C(M) = \{X \in \mathcal{M} \ : \ XM = MX\}$ and $C(M)^{\perp} = \{i[M, X] \ : \ X \in \mathcal{M}\}$.

Let $H \in \tilde{T}_M$ be decomposed as $H = H^c + i[M, X]$, then, according to [4] (p. 124), $\frac{d}{ds} g(M + sH)|_{s=0} = g'(M)H^c + i[g(M), X]$. The statements (i) and (ii) follow easily from this equality. Let $K \in \tilde{T}_M$, $K = K^c + i[M, Y]$, then

$$\begin{aligned}\operatorname{Tr} K L_g[M](H) &= \operatorname{Tr} K^c g'(M) H^c - \operatorname{Tr}[M, Y][g(M), X] \\ &= \operatorname{Tr} g'(M) K^c H^c - \operatorname{Tr}[g(M), Y][M, X] = \operatorname{Tr} L_g[M](K) H.\end{aligned}$$

which proves (iii).

(iv) $L_g[M](M) = \frac{d}{ds} g((1+s)M)|_{s=0} = g'(M)M$.

(v) From (ii),

$$L_g[M]^{-1}(I) = L_{g^{-1}}[g(M)](I) = \frac{d}{ds} g^{-1}(g(M) + sI)|_{s=0} = (g'(M))^{-1}. \quad \square$$

In what follows, we omit the indication of the point in square brackets if no confusion is possible. The vector space

$$T_D^g = \{L_g(H) \mid H \in T_D\}$$

will be called the $g$-representation of the tangent space $T_D$. The corresponding inner product in $T_D^g$ is

$$\lambda_D^g(G_1, G_2) = \lambda_D(L_g^{-1}(G_1), L_g^{-1}(G_2)) = \operatorname{Tr} G_1 K_g(G_2),$$

where $K_g = L_g^{-1} J_D L_g^{-1}$. Similarly as before, the $g$-representation of the cotangent space is the space of all linear functionals on $T_D^g$ and it is given by

$$T_D^{g*} = \{K_g(G) \mid G \in T_D^g\} = \{L_{g^{-1}}(A) \mid A \in T_D^*\}.$$

The inner product

$$\varphi_D^g(B_1, B_2) = \lambda_D^g(K_g^{-1}(B_1), K_g^{-1}(B_2)) = \operatorname{Tr} B_1 K_g^{-1}(B_2)$$

will be called the generalized $g$-covariance of $B_1$ and $B_2$.

Clearly, if $g$ is the identity function, we obtain the usual tangent and cotangent spaces $T_D$ and $T_D^*$.

LEMMA 3.2.

$$\begin{aligned} G \in T_D^g &\iff \operatorname{Tr}(g'(D))^{-1} G = 0, \\ B \in T_D^{g*} &\iff \operatorname{Tr} g'(D) D B = 0. \end{aligned}$$

*Proof:* Both statements follow easily from $\operatorname{Tr} H = 0$, $H \in T_D$ and Lemma 3.1(v) and (iv), respectively. $\square$

EXAMPLE 3.1. The quantum analogue of Amari's $\alpha$-representations is obtained if we put

$$g(x) = g_\alpha(x).$$

In the sequel, we use the letter $\alpha$ to indicate the function $g_\alpha$, e.g. $\alpha$-representation, $T_D^\alpha$, etc. For $\alpha \neq \pm 1$, $(g'_\alpha(D))^{-1} = \frac{1+\alpha}{2} g_{-\alpha}(D)$ and $g'_\alpha(D)D = \frac{1-\alpha}{2} g_\alpha(D)$ and thus

$$G \in T_D^\alpha \iff \operatorname{Tr} g_{-\alpha}(D)G = 0,$$
$$B \in T_D^{\alpha *} \iff \operatorname{Tr} g_\alpha(D)B = 0.$$

An application of Lemma 3.2 also for $\alpha = \pm 1$ shows that $T_D^{\alpha *} = T_D^{-\alpha}$ for each $\alpha$, so that $K_\alpha = L_\alpha^{-1} J_D L_\alpha^{-1}$ is an isomorphism $K_\alpha : T_D^\alpha \to T_D^{-\alpha}$. This shows that the $\alpha$- and $-\alpha$-representations are in some sense dual, as in the classical case. In particular, if we put $J_D = J_\alpha = L_{-\alpha} L_\alpha$, then $K_\alpha = L_{-\alpha} L_\alpha^{-1}$ and we see that in this case $K_\alpha^{-1} = K_{-\alpha}$. The corresponding family of metrics was studied in [10] and it was shown that the metric is monotone for $\alpha \in [-3, 3]$. Moreover, for $\alpha = \pm 1$ we obtain the Kubo–Mori metric and $\alpha = \pm 3$ corresponds to the right logarithmic derivative.

## 4. The affine connections, torsion and curvature

Let $g : \mathbb{R} \to \mathbb{R}$ be a smooth strictly monotone function and let $M, M' \in \mathcal{M}^+$. Clearly, both $T_M^g$ and $T_M^{g*}$ for each $M$ can be identified with $\mathcal{M}$, so that there is a natural isomorphism $\tilde{T}_M^g \to \tilde{T}_{M'}^g$, given by the identity mapping. This isomorphism induces an affine connection on $\mathcal{M}^+$. Let us denote the corresponding covariant derivative by $\tilde{\nabla}^g$.

Let $x_1, \ldots, x_{N+1}$ be a coordinate system in $\mathcal{M}^+$. Let us denote $\partial_i = \frac{\partial}{\partial x_i}$ and let $\tilde{H}_i = \partial_i M(x)$, $\tilde{G}_i = L_g(\tilde{H}_i) = \partial_i g(M(x))$, $i = 1, \ldots, N+1$. Then

$$L_g(\tilde{\nabla}^g_{\tilde{H}_j} \tilde{H}_i) = \partial_i \partial_j g(M(x)).$$

Hence, the coefficients of the affine connection are

$$\tilde{\Gamma}^g_{ijk}(x) = \lambda_x(\tilde{\nabla}^g_{\tilde{H}_i} \tilde{H}_j, \tilde{H}_k) = \operatorname{Tr} \partial_i \partial_j g(M(x)) K_g(\tilde{G}_k),$$

$i, j, k = 1, \ldots, N+1$. From this, it follows that this connection is torsion-free.

If we use the functions $g_\alpha$, we obtain a one-parameter family of torsion-free connections $\tilde{\nabla}^\alpha$, analogical to Amari's family of $\alpha$-connections. It is easy to see that, unlike in the classical case, the connections $\tilde{\nabla}^\alpha$ and $\tilde{\nabla}^{-\alpha}$ are not dual in general.

To obtain the dual connection, consider the affine connection on $\mathcal{M}^+$ induced by a similar identification $\tilde{T}_M^{g*} \to \tilde{T}_{M'}^{g*}$. The covariant derivative will be denoted by $\tilde{\nabla}^{g*}$, this notation will be justified below. We have

$$L_g^{-1} J_M(\tilde{\nabla}^{g*}_{\tilde{H}_i} \tilde{H}_j) = \partial_i K_g(\tilde{G}_j(x)).$$

The coefficients of this connection are

$$\tilde{\Gamma}^{g*}_{ijk} = \lambda(\tilde{\nabla}^{g*}_{\tilde{H}_i} \tilde{H}_j, \tilde{H}_k) = \varphi^g(L_g^{-1} J_M(\tilde{\nabla}^{g*}_{\tilde{H}_i} \tilde{H}_j), K_g(\tilde{G}_k)) = \operatorname{Tr} \partial_i K_g(\tilde{G}_j) \tilde{G}_k. \quad (5)$$

PROPOSITION 4.1. *$\tilde{\nabla}^g$ and $\tilde{\nabla}^{g*}$ are dual.*

*Proof:* For $i, j, k = 1, \ldots, N+1$, we have

$$\partial_i \lambda(\tilde{H}_j, \tilde{H}_k) = \partial_i \lambda^g(\tilde{G}_j, \tilde{G}_k) = \partial_i \mathrm{Tr}\, \tilde{G}_j K_g(\tilde{G}_k)$$
$$= \mathrm{Tr}\, \partial_i \tilde{G}_j K_g(\tilde{G}_k) + \mathrm{Tr}\, \tilde{G}_j \partial_i K_g(\tilde{G}_k) = \tilde{\Gamma}^g_{ijk} + \tilde{\Gamma}^{g*}_{ikj}.$$

REMARK 4.1. Notice that for $g = \mathrm{id}$, we obtain the mixture and exponential $\nabla^{(m)}$ and $\nabla^{(e)}$ connections defined in [13].

The components of the torsion tensor are

$$\tilde{S}^{g*}_{ijk} = \tilde{\Gamma}^{g*}_{ijk} - \tilde{\Gamma}^{g*}_{jik} = \mathrm{Tr}\, \{\partial_i K_g(\tilde{G}_j(x)) - \partial_j K_g(\tilde{G}_i(x))\}\tilde{G}_k,$$

so that this connection is torsion-free iff

$$\partial_i K_g(\tilde{G}_j) = \partial_j K_g(\tilde{G}_i), \quad \forall i, j = 1, \ldots, N+1.$$

Obviously, this is not always the case.

EXAMPLE 4.1. Let us consider the connection $\tilde{\nabla}^{\alpha*}$, $\alpha \in [-3, 3]$, and let the metric tensor be determined by $J_\alpha$. Then we have $K_\alpha(\tilde{G}_j(x)) = L_{-\alpha}(\tilde{H}_j(x)) = \partial_j g_{-\alpha}(M(x))$, so that the connection is torsion-free. Moreover, $L_\alpha^{-1} J_\alpha = L_{-\alpha}$ and thus

$$L_{-\alpha}(\tilde{\nabla}^{\alpha*}_{\tilde{H}_i} \tilde{H}_j) = \partial_i \partial_j g_{-\alpha}(M(x)).$$

From this it follows that with this choice of the metric tensor one has $\tilde{\nabla}^{\alpha*} = \tilde{\nabla}^{-\alpha}$ as in the classical case. In particular, the exponential connection $\tilde{\nabla}^{-1*}$ is torsion-free and $\tilde{\nabla}^{\pm 1*} = \tilde{\nabla}^{\mp 1}$ with the Kubo–Mori metric. Similarly, $\tilde{\nabla}^{\pm 3*} = \tilde{\nabla}^{\mp 3}$ and the connections are torsion-free, with the metric of the right logarithmic derivative.

Consider now $\mathcal{D}$ as an $N$-dimensional submanifold in $\mathcal{M}$ and let $t_1, \ldots, t_N$ be a coordinate system in $\mathcal{D}$. As there is no danger of confusion, we use the symbol $\partial_i$ also for $\frac{\partial}{\partial t_i}$. Let $H_i = \partial_i D(t)$, $G_i = \partial_i g(D(t))$, $i = 1, \ldots, N$. The affine structure in $\mathcal{D}$ is obtained by projecting the above affine connections orthogonaly onto $\mathcal{D}$. Clearly, each density matrix $D$ is orthogonal to the tangent space $T_D$ in $\tilde{T}_D$. Indeed, if $H \in T_D$,

$$\lambda_D(H, D) = \mathrm{Tr}\, H J_D(D) = \mathrm{Tr}\, H = 0.$$

Moreover, $\lambda_D(D, D) = 1$. Using (iv) and (v) of Lemma 3.1, it follows that the covariant derivative is given by

$$L_g(\nabla^g_{H_j} H_i) = \partial_i \partial_j g(D_t) - g'(D_t) D_t \mathrm{Tr}\, (g'(D_t))^{-1} \partial_i \partial_j g(D_t),$$
$$L_g^{-1} J_D(\nabla^{g*}_{H_i} H_j) = \partial_i K_g(G_j(t)) - (g'(D_t))^{-1} \mathrm{Tr}\, g'(D_t) D_t \partial_i K_g(G_j(t)),$$

and the coefficients are

$$\Gamma^{g}_{ijk}(t) = \operatorname{Tr} K_g(G_k)\partial_i\partial_j g(D_t),$$
$$\Gamma^{g*}_{ijk}(t) = \operatorname{Tr} G_k \partial_i K_g(G_j(t)).$$

And now some facts from differential geometry [1, 2]. Let $\tilde{R}$ be the Riemannian curvature tensor of an affine connection $\tilde{\nabla}$ on an $m$-dimensional manifold $\mathcal{M}$ and let $\tilde{R}^*$ be the curvature tensor of the dual connection. Let $X$, $Y$, $Z$, $W$ be vector fields. Then we have (see Lauritzen in [2])

$$\tilde{R}(X, Y, Z, W) = -\tilde{R}^*(X, Y, W, Z). \tag{6}$$

In particular, $\tilde{R} = 0$ iff $\tilde{R}^* = 0$. Further, let $\mathcal{N}$ be a $p$-dimensional submanifold in $\mathcal{M}$ with a coordinate system $x$. Let $X_1, \ldots, X_p$ be the natural basis of the tangent space, associated with $x$, and let $Y_1, \ldots, Y_{m-p}$ be orthonormal vector fields on $\mathcal{M}$ normal to $\mathcal{N}$. Recall that the Euler–Schouten imbedding curvature is given by

$$H_{ijl} = \lambda(\tilde{\nabla}_{X_i} X_j, Y_l), \quad i, j = 1, \ldots, p, \;\; l = 1, \ldots, m-p.$$

Let us denote

$$H^*_{ijl} = \lambda(\tilde{\nabla}^*_{X_i} X_j, Y_l), \quad i, j = 1, \ldots, p, \;\; l = 1, \ldots, m-p.$$

The submanifold $\mathcal{N}$ is called autoparallel if its imbedding curvature vanishes, i.e. the parallel shift of a vector in $T_x(\mathcal{N})$ along a curve $\rho(s)$ in $\mathcal{N}$ stays in $T_{\rho(s)}(\mathcal{N})$. Let $\nabla$ be the orthogonal projection of $\tilde{\nabla}$ onto $\mathcal{N}$.

PROPOSITION 4.2. *Let $R$ be the Riemannian curvature tensor of $\nabla$ and let $\tilde{\nabla}$ be torsion-free. Then*

$$\tilde{R}_{ijkl} = R_{ijkl} + \sum_{\nu}(H_{ik\nu}H^*_{jl\nu} - H_{jk\nu}H^*_{il\nu}) \tag{7}$$

*for $i, j, k, l = 1, \ldots, p$.*

*Proof:* The proof of this statement for the case of a metric connection, i.e. $\nabla = \nabla^*$, can be found in [12], the general case is obtained by an easy modification of this proof.

Let us now return to the submanifold $\mathcal{D}$ in $\mathcal{M}^+$. The Euler–Schouten imbedding curvature is

$$H^{g}_{ij1} = \operatorname{Tr}(g'(D))^{-1}\partial_i\partial_j g(D), \quad i, j = 1, \ldots, N,$$

and

$$H^{g*}_{ij1} = \operatorname{Tr} g'(D) D \partial_i K_g(G_j), \quad i, j = 1, \ldots, N.$$

PROPOSITION 4.3. *Let $g = g_\alpha$. Then*

$$H^{\alpha}_{ij1} = -\frac{1+\alpha}{2}\operatorname{Tr} H_i J_\alpha(H_j),$$
$$H^{\alpha*}_{ij1} = -\frac{1-\alpha}{2}\operatorname{Tr} H_i J_D(H_j).$$

*Proof:* Let $\alpha \neq \pm 1$. Compute, using Lemma 3.2,

$$\begin{aligned} H^{\alpha}_{ij1} &= \frac{1+\alpha}{2} \mathrm{Tr}\, g_{-\alpha}(D) \partial_i \partial_j g_{\alpha}(D) \\ &= \frac{1+\alpha}{2} \{\partial_j \mathrm{Tr}\, g_{-\alpha}(D) \partial_i g_{\alpha}(D) - \mathrm{Tr}\, \partial_j g_{-\alpha}(D) \partial_i g_{\alpha}(D)\} \\ &= -\frac{1+\alpha}{2} \mathrm{Tr}\, L_{-\alpha}(H_j) L_{\alpha}(H_i) \end{aligned}$$

and

$$\begin{aligned} H^{\alpha *}_{ij1} &= \frac{1-\alpha}{2} \mathrm{Tr}\, g_{\alpha}(D) \partial_i K_{\alpha}(G_j) \\ &= \frac{1-\alpha}{2} \{\partial_i \mathrm{Tr}\, g_{\alpha}(D) L_{\alpha}^{-1} J_D(H_j) - \mathrm{Tr}\, \partial_i g_{\alpha}(D) L_{\alpha}^{-1} J_D(H_j)\} \\ &= -\frac{1-\alpha}{2} \mathrm{Tr}\, H_j J_D(H_i). \end{aligned}$$

The proof in the case $\alpha = \pm 1$ is almost the same. □

Let $\tilde{R}^g$ be the Riemannian curvature tensor of $\tilde{\nabla}^g$ in $\mathcal{M}^+$. Clearly, $\mathcal{M}^+$ can be parametrized in such a way that

$$g(M(x)) = \sum_{i=1}^{N+1} x_i G_i. \tag{8}$$

In this case, we will say that $\mathcal{M}^+$ is parametrized as an extended $g$-family. We have $\partial_i \partial_j g(M(x)) = 0$, hence $\tilde{\Gamma}^g_{ijk}(x) = 0$, for $i, j, k = 1, \ldots, N+1$ and for each $x$. It means that this parametrization is affine and therefore $\tilde{R}^g = 0$. Thus also $\tilde{R}^{g*} = 0$. From the identity (7), the curvature tensor $R^g$ in $\mathcal{D}$ is equal to

$$R^g_{ijkl} = H^g_{jk1} H^{g*}_{il1} - H^g_{ik1} H^{g*}_{jl1}.$$

If $g = g_{\alpha}$,

$$R^{\alpha}_{ijkl}(D) = \frac{1-\alpha^2}{4} \{\mathrm{Tr}\, H_j J_{\alpha}(H_k) \mathrm{Tr}\, H_i J_D(H_l) - \mathrm{Tr}\, H_i J_{\alpha}(H_k) \mathrm{Tr}\, H_j J_D(H_l)\}.$$

We see that $R^{\alpha} = 0$ if $\alpha = \pm 1$. It is also clear that $R^{\alpha} = R^{-\alpha}$. Statistical manifolds with this property are called conjugate symmetric and were studied by Lauritzen (see [2]). Here we have to be aware that $R^{-\alpha} \neq R^{\alpha *}$. The curvature tensor $R^{\alpha *}$ can be computed using the identity (6).

Let $\mathcal{M}'$ be a submanifold in $\mathcal{M}^+$ such that the Riemannian curvature tensor of the projection $\nabla^{\alpha'}$ of $\tilde{\nabla}^{\alpha}$ onto $\mathcal{M}'$ vanishes. Then $\mathcal{M}'$ is flat so that there exists an affine coordinate system $\theta$. The dual connection is curvature-free and if it is also torsion-free then it follows from Theorem 1.2 that there exists a dual $\nabla^{\alpha'*}$-affine coordinate system $\eta$. This is the case if, for example, $J = J_\alpha$ or if $\mathcal{M}'$ is one-dimensional, see Section 6. However, if the dual connection is not torsion-free, there is no coordinate system dual to $\theta$. Indeed, if $\eta$ is a dual coordinate system, then it follows from Amari's proof of Theorem 1.2 that $\eta$ is $\nabla^{\alpha'*}$-affine, i.e. the components of the connection vanish, $\Gamma^{\alpha*}_{ijk}(\eta) = 0$ $\forall\eta$, $\forall i, j, k$. But then we have for the components of the torsion tensor $S^{\alpha*}_{ijk} = \Gamma^{\alpha*}_{ijk} - \Gamma^{\alpha*}_{jik} = 0$.

## 5. A statistical interpretation

Throughout this paragraph we suppose that $\alpha \neq 1$. We will investigate a statistical interpretation of the $\alpha$-representations of the tangent and cotangent space and the $\alpha$-connections.

Let $\mathcal{D}' \subseteq \mathcal{D}$ be a smooth $p$-dimensional submanifold and let $\theta_1, \dots, \theta_p$ be the coordinate system in $\mathcal{D}'$. Let $T'_\theta$ be the tangent space of $\mathcal{D}'$ at $\theta$, $H'_i = \frac{\partial}{\partial\theta_i}D(\theta)$, $G'_i = L_g(H'_i)$ and let $\nabla'^g$ and $\nabla'^{g*}$ denote the orthogonal projections of the affine connections onto $\mathcal{D}'$. In [17], locally unbiased estimators were defined and a generalized Cramèr–Rao inequality was proved for the generalized covariance. We give an analogical definition of the $\alpha$-expectation and $\alpha$-unbiasedness.

DEFINITION 5.1. Let $B = (B_1, \dots, B_p)$ be a collection of observables. We will say that $B$ is a *locally $\alpha$-unbiased estimator of $\theta$ at $\theta_0$* if

(i) $\operatorname{Tr} g_\alpha(D_{\theta_0})B_i = \theta_{0i}$ for $i = 1, \dots, p$,

(ii) $\frac{\partial}{\partial\theta_j}\operatorname{Tr} g_\alpha(D_\theta)B_i|_{\theta_0} = \operatorname{Tr} G'_jB_i = \delta_{ij}$, $i, j, = 1, \dots, p$.

The value $\operatorname{Tr} g_\alpha(D)A$ will be called the *$\alpha$-expectation of the observable $A$*.

As follows from Example 3.1, the $\alpha$-representation of the cotangent space $T^{\alpha*}_D$ can be interpreted as the space of all observables with zero $\alpha$-expectation at $D$ with inner product given by the generalized $\alpha$-covariance $\varphi^\alpha$. The following lemma is obvious.

LEMMA 5.1. *Let $A = (A_1, \dots, A_p)$ be a locally unbiased estimator of $\theta$ at $\theta_0$. Then $B = L_\alpha^{-1}(A) = (L_\alpha^{-1}(A_1), \dots, L_\alpha^{-1}(A_p))$ is a locally $\alpha$-unbiased estimator of $\theta$ at $\theta_0$. Moreover, $\varphi(A_i, A_j) = \varphi^\alpha(B_i, B_j)$ and $\operatorname{Tr} H'_iA_j = \operatorname{Tr} G'_iB_j$, $i, j = 1, \dots, p$.*

The generalized Cramér–Rao inequality from [17] can now be rewritten in the following form.

THEOREM 5.1. *Let $\lambda_{ij} = \lambda(H'_i, H'_j) = \lambda^\alpha(G'_i, G'_j)$ and let $B = (B_1, \dots, B_p)$ be a locally $\alpha$-unbiased estimator of $\theta$ at 0. Then*

$$\varphi^\alpha_D(B) \geq (\lambda_{ij})^{-1}$$

*in the sense of the order on positive definite matrices. Moreover, equality is attained iff $B$ is the biorthogonal basis of $T'^{\alpha*}_0$.*

An estimator $B$ which is ($\alpha$-)unbiased (at each point) and such that its variance attains the Cramèr–Rao bound is called ($\alpha$-)efficient. Clearly, such estimator does not always exist. The following necessary and sufficient condition is a generalization of a result stated (without proof) in [13].

THEOREM 5.2. *The $\alpha$-efficient estimator exists iff $\mathcal{D}'$ is a $\nabla^{\alpha*}$-autoparallel submanifold and the coordinate system $\theta_1, \ldots, \theta_p$ is $\nabla'^{\alpha}$-affine.*

*Proof:* Let $\mathcal{D}'$ be $\nabla^{\alpha*}$-autoparallel and let the coordinate system be $\nabla'^{\alpha}$-affine, i.e. $\nabla'^{\alpha}_{H'_i} H'_j = 0$, $i, j = 1, \ldots, p$. Let us choose a point in the parameter space ($\theta = 0$) and let $B = (B_1, \ldots, B_p)$ be the biorthogonal basis of $T_0'^{\alpha*}$. We prove that $B$ is an $\alpha$-efficient estimator.

Let $X_i(\theta)$ be a $\nabla^{\alpha*}$-parallel vector field such that $L_\alpha^{-1} J_0(X_i(0)) = B_i$, $i = 1, \ldots, p$, i.e. $X_i = J_\theta^{-1} L_\alpha \{B_i - g_{-\alpha}(D(\theta)) \mathrm{Tr}\, g_\alpha(D(\theta)) B_i\}$. As $\mathcal{D}'$ is $\nabla^{\alpha*}$-autoparallel, $X_j(\theta) \in T'_\theta$, $\forall \theta$. Compute

$$\partial_i \mathrm{Tr}\, g_\alpha(D(\theta)) B_j = \mathrm{Tr}\, G'_i(\theta) B_j = \mathrm{Tr}\, G'_i(\theta) L_\alpha^{-1} J_\theta(X_j(\theta)) = \lambda(H'_i, X_j),$$

here we have used the identity $\mathrm{Tr}\, G'_i(\theta) g_{-\alpha}(D(\theta)) = 0$. Further,

$$\partial_k \lambda(H'_i, X_j) = \lambda(\nabla'^{\alpha}_{H'_k} H'_i, X_j) + \lambda(H'_i, \nabla'^{\alpha*}_{H'_k} X_j). \tag{9}$$

Since the parametrization is $\nabla'^{\alpha}$-affine, we have $\nabla'^{\alpha}_{H'_k} H'_i = 0$, $\forall i, k$. Moreover, $X_j(\theta) \in T'_\theta$ and $X_j$ is $\nabla^{\alpha*}$-parallel, hence $X_j$ is $\nabla'^{\alpha*}$-parallel, so that $\nabla'^{\alpha*}_{H'_i} X_j = 0$. It follows that $\partial_k \mathrm{Tr}\, G'_i(\theta) B_j = 0$ for all $\theta$. Since $\mathrm{Tr}\, G'_i(0) B_j = \delta_{ij}$, we have $\mathrm{Tr}\, G'_i(\theta) B_j = \delta_{ij}$ for each $\theta$. We see that as $\mathrm{Tr}\, g_\alpha(D(0)) B_j = 0$ and $\partial_i \mathrm{Tr}\, g_\alpha(D(\theta)) B_j = \delta_{ij}$ for each $\theta$, it follows that $B$ is $\alpha$-unbiased at each point $\theta$. From Theorem 5.1, it now suffices to prove that $B_j - \theta_j g_{-\alpha}(D(\theta)) \in T_\theta'^{\alpha*}$. But this follows easily from the fact that $X_j(\theta) \in T'_\theta$ and $\mathrm{Tr}\, g_\alpha(D(\theta)) B_j = \theta_j$.

Conversely, let $B$ be the $\alpha$-efficient estimator, then $B$ is $\alpha$-unbiased and the matrices $B_i - \theta_i g_{-\alpha}(D(\theta))$, $i = 1, \ldots, p$, form the biorthogonal basis of $T_\theta'^{\alpha*}$ $\forall \theta$. Let $X_i = J_\theta^{-1} L_\alpha^{-1}(B_i - \theta_i g_{-\alpha}(D(\theta)))$, then $X_i$ is a $\nabla^{\alpha*}$-parallel vector field and $X_i \in T'_\theta$ $\forall \theta$, hence $X_i$ is $\nabla'^{\alpha*}$-parallel. Moreover,

$$\lambda(H'_i, X_j) = \mathrm{Tr}\, G'_i B_j = \delta_{ij}.$$

From (9) it now follows that $\lambda(\nabla'^{\alpha}_{H'_k} H'_i, X_j) = 0$. But the matrices $X_j(\theta)$, $j = 1, \ldots, p$, form a basis of $T'_\theta$. We may conclude that $\nabla'^{\alpha}_{H'_k} H'_i = 0$, $i, k = 1, \ldots, p$, so that the parametrization is $\nabla'^{\alpha}$-affine.

To see that $\mathcal{D}'$ is $\nabla^{\alpha*}$-autoparallel, it suffices to observe that the parallel vector fields $X_i(\theta)$, $i = 1, \ldots, p$, form a basis of the tangent space $T'_\theta$ for each $\theta$. □

## 6. Geodesics and divergence functions

Let us consider a $\nabla^\alpha$-autoparallel submanifold in $\mathcal{D}$ for $\alpha = \pm 1$. Then it is $\nabla^\alpha$-flat, hence there is an affine coordinate system $\theta$. If $\alpha \neq \pm 1$, we will consider the autoparallel submanifolds in $\mathcal{M}^+$. As it was said at the end of Section 4, in general there is no hope for a dual coordinate system to exist, unless the dual connection is torsion-free. It means that we cannot use Amari's theory to define a divergence. However, one-dimensional submanifolds are always torsion-free, so that a divergence function exists for each $\nabla^{\pm 1*}$ and $\tilde{\nabla}^{\alpha*}$-geodesic. As suggested in [13], we use these functions to define a divergence function in $\mathcal{D}$ ($\mathcal{M}^+$).

Clearly, for each $\alpha$, a $\tilde{\nabla}^{\alpha*}$-geodesic is a solution of

$$L_\alpha^{-1} J_{\rho_t}(\dot{\rho}_t) = A, \tag{10}$$

where $A \in \mathcal{M}$, it means that each geodesic is determined by the observable $A$.

The $\nabla^{\alpha*}$-geodesic is given by

(i) $J_{\rho_t}(\dot{\rho}_t) = A - \operatorname{Tr} \rho_t A, \quad$ for $\alpha = -1$,

(ii) $L_1^{-1} J_{\rho_t}(\dot{\rho}_t) = A - \rho_t \operatorname{Tr} A, \quad$ for $\alpha = 1$,

(iii) $L_\alpha^{-1} J_{\rho_t}(\dot{\rho}_t) = A - \frac{1-\alpha^2}{4} g_{-\alpha}(\rho_t) \operatorname{Tr} g_\alpha(\rho_t) A, \quad$ for $\alpha \neq \pm 1$.

Note that $A$ can be replaced by $A + c_t$ in (i), $A + c_t \rho_t$ in (ii) and $A + c_t g_{-\alpha}(\rho_t)$ in (iii), $c_t \in R$, so that we may always suppose that $A \in T_t^{\alpha*}$.

The relation between $\nabla^{\alpha*}$- and $\tilde{\nabla}^{\alpha*}$-geodesics is clarified in the following proposition.

PROPOSITION 6.1. *Let $\tilde{\rho}_t$ be a solution of $L_\alpha^{-1} J_{\rho_t}(\dot{\rho}_t) = A$. Then $\rho_t = \frac{\tilde{\rho}_t}{\operatorname{Tr} \tilde{\rho}_t}$ is a $\nabla^{\alpha*}$-geodesic.*

*Proof:* From (3), $J_{\rho_t} = \operatorname{Tr} \tilde{\rho}_t J_{\tilde{\rho}_t}$. Compute

$$J_{\rho_t}(\dot{\rho}_t) = J_{\tilde{\rho}_t}(\dot{\tilde{\rho}}_t) - \frac{\operatorname{Tr} \dot{\tilde{\rho}}_t}{\operatorname{Tr} \tilde{\rho}_t} = J_{\tilde{\rho}_t}(\dot{\tilde{\rho}}_t) - \operatorname{Tr} \rho_t J_{\tilde{\rho}_t}(\dot{\tilde{\rho}}_t);$$

here we used the fact that $J_D(D) = I$ (twice) and that $J_D$ is self-adjoint. Further,

$$L_\alpha^{-1} J_{\rho_t}(\dot{\rho}_t) = A - L_\alpha^{-1}(I) \operatorname{Tr} L_\alpha(\rho_t) A.$$

We use Lemma 3.1 (iv) and (v) to complete the proof. □

REMARK 6.1. Let $\rho_t$ be as in (i). For each $t$, the coefficient of the affine connection is equal to

$$\Gamma^{-1*}(t) = \lambda(\nabla_{\dot{\rho}_t}^{-1*} \dot{\rho}_t, \dot{\rho}_t) = \operatorname{Tr} \frac{d}{dt} \{J_{\rho_t}(\dot{\rho}_t)\} \dot{\rho}_t = \operatorname{Tr} \frac{d}{dt} \{A - \operatorname{Tr} \rho_t A\} \dot{\rho}_t = 0.$$

It follows that the parameter $t$ is $\nabla^{-1*}$-affine. Similarly, for $\rho_t$ as in (ii), $t$ is $\nabla^{1*}$-affine. In the case $\alpha \neq \pm 1$, the $\nabla^{\alpha*}$-geodesic is not flat, hence we consider (as in the classical case) geodesics in $\mathcal{M}^+$. If $\rho_t$ is the solution of (10), $t$ is $\tilde{\nabla}^{\alpha *}$-affine. Moreover, all affine coordinate systems are connected via affine transformations $t \mapsto at + b$. This coordinate transformation changes the initial point and rescales the observable $A$ as $\frac{1}{a}A$.

LEMMA 6.1. *Let $\alpha = -1$ and let $\rho_t$ and $\tilde{\rho}_t$ be as above. Then $\psi(t) = \log \operatorname{Tr} \tilde{\rho}_t$ is a potential function for $\rho_t$.*

*Proof:* We have to prove that $\frac{d^2}{dt^2}\psi(t) = \lambda(\dot{\rho}_t, \dot{\rho}_t)$. Compute

$$\frac{d}{dt}\psi(t) = \frac{\operatorname{Tr} \dot{\tilde{\rho}}_t}{\operatorname{Tr} \tilde{\rho}_t} = \operatorname{Tr} \rho_t A.$$

Moreover,

$$\lambda(\dot{\rho}_t, \dot{\rho}_t) = \operatorname{Tr} \dot{\rho}_t J_{\rho_t}(\dot{\rho}_t) = \operatorname{Tr} \dot{\rho}_t A = \frac{d^2}{dt^2}\psi(t). \qquad \square$$

LEMMA 6.2. *Let $\alpha = 1$, then $\psi(t) = \operatorname{Tr} \rho_t \log \rho_t$ is a potential function for $\rho_t$.*

*Proof:* We may assume that $A \in T_{\rho_0}^{1*}$, i.e. $\operatorname{Tr} A = 0$. We have $L_1^{-1}(\log \rho_t) = \rho_t \log \rho_t$, hence

$$\psi(t) = \operatorname{Tr} L_1^{-1}(\log \rho_t) = \operatorname{Tr} \rho_t J_{\rho_t} L_1^{-1}(\log \rho_t)$$

and

$$\begin{aligned} \frac{d}{dt}\psi(t) &= \operatorname{Tr} \dot{\rho}_t J_{\rho_t} L_1^{-1}(\log \rho_t) + \operatorname{Tr} \rho_t \frac{d}{dt} J_{\rho_t} L_1^{-1}(\log \rho_t) \\ &= \operatorname{Tr} A \log \rho_t + \operatorname{Tr} \rho_t L_1(\dot{\rho}_t) = \operatorname{Tr} A \log \rho_t; \end{aligned}$$

here we have used the fact that $J_D(A) = D^{-1}A$ whenever $A$ commutes with $D$, thus $J_{\rho_t}(\rho_t \log \rho_t) = \log \rho_t$, and $L_1(\rho_t) = I$. The rest of the proof is the same as above. $\square$

LEMMA 6.3. *Let $\alpha \neq \pm 1$ and let $\rho_t = \tilde{\rho}_t$. Then $\psi(t) = \frac{2}{1-\alpha} \operatorname{Tr} \rho_t$ is the potential function .*

*Proof:*

$$\frac{d}{dt}\operatorname{Tr} \rho_t = \operatorname{Tr} \dot{\rho}_t J_{\rho_t}(\rho_t) = \operatorname{Tr} \rho_t L_\alpha(A) = \frac{1-\alpha}{2}\operatorname{Tr} g_\alpha(\rho_t) A$$

and

$$\frac{d^2}{dt^2}\operatorname{Tr} \rho_t = \frac{1-\alpha}{2}\operatorname{Tr} L_\alpha(\dot{\rho}_t) A = \frac{1-\alpha}{2}\lambda(\dot{\rho}_t, \dot{\rho}_t). \qquad \square$$

According to Theorem 1.2, $\rho_t$ is also $\nabla^\alpha$ ($\tilde{\nabla}^\alpha$)-flat and there is a dual $\nabla^\alpha$ ($\tilde{\nabla}^\alpha$)-affine coordinate $s$. As we have seen in the proofs of the above lemmas, the dual coordinate is given by $s(t) = \frac{d}{dt}\psi(t) = \operatorname{Tr} g_\alpha(\rho_t) A$ for each $\alpha$. Moreover, there is a divergence function $D_\alpha^\rho : \rho \times \rho \to \mathbb{R}$. A divergence measure in $\mathcal{D}$ ($\mathcal{M}^+$) can then be defined as follows. Let $\rho_0, \rho_1 \in \mathcal{D}$. Let $\tilde{\rho}_t$ be the unique $\tilde{\nabla}^{\alpha *}$-geodesic connecting these two states. If $\alpha \neq \pm 1$, the $\alpha$-divergence is $D_\alpha(\rho_0, \rho_1) = D_\alpha^{\tilde{\rho}}(0, 1)$. If $\alpha = \pm 1$, we use the $\nabla^{\alpha *}$-geodesic $\rho_t = (\operatorname{Tr} \tilde{\rho}_t)^{-1} \tilde{\rho}_t$.

PROPOSITION 6.2.

(i) *Let $\alpha = -1$ and let $\rho_0, \rho_1 \in \mathcal{D}$. Let $\rho_t$ be as above and let $A$ be the unique observable determining $\rho_t$. Then*

$$D_\alpha(\rho_0, \rho_1) = \operatorname{Tr} \rho_1 A.$$

(ii) *If $\alpha = 1$, then*

$$D_\alpha(\rho_0, \rho_1) = \operatorname{Tr} \rho_0 \log \rho_0 + \operatorname{Tr} (A - \rho_1) \log \rho_1.$$

(iii) *Let $\alpha \neq \pm 1$, $\rho_0, \rho_1 \in \mathcal{M}^+$. Then*

$$D_\alpha(\rho_0, \rho_1) = \frac{2}{1-\alpha}(\operatorname{Tr} \rho_0 - \operatorname{Tr} \rho_1) + \operatorname{Tr} g_\alpha(\rho_1) A.$$

*Proof:* From the definition of the divergence function and the identity (2), we obtain

$$D_\alpha^\rho(t_1, t_2) = \psi(t_1) - \psi(t_2) + (t_2 - t_1) s_2$$

The rest of the proof is easy. □

Let now $\rho_t$ be a geodesic connecting two states $\rho_0$ and $\rho_1$ and let $\rho_{t_1}, \rho_{t_2}$ be two states lying on $\rho_t$. Using Remark 6.1, it is easy to see that $D_\alpha(\rho_{t_1}, \rho_{t_2}) = D_\alpha^\rho(t_1, t_2)$. It also follows that for each $\rho$ we may put $D_\alpha(\rho, \rho) = 0$. There are some properties of the divergence $D_\alpha$ which follow from the properties of $D_\alpha^\rho$.

(i) Positivity: $D_\alpha(\rho, \sigma) \geq 0$ and $D_\alpha(\rho, \sigma) = 0$ iff $\rho = \sigma$.

(ii) Let $\sigma = \rho + dtH$. Let $A = L_\alpha^{-1} J_D(H)$ and let $\rho_t$ be the geodesic determined by $A$ such that $\rho_{t_0} = \rho$. Then

$$D_\alpha(\rho, \rho + dtH) = D_\alpha^\rho(t_0, t_0 + dt) = dt^2 \lambda(H, H)$$

and

$$D_\alpha(\rho + dtH, \rho) = D_\alpha^\rho(t_0 + dt, t_0) = D_\alpha^\rho(t_0, t_0 + dt).$$

It means that the $\alpha$-divergence induces the metric.

EXAMPLE 6.1. Let $\lambda$ be determined by $J_\alpha = L_{-\alpha} L_\alpha$ for some $\alpha \in (-3, 3)$. As we have seen, in this case $\nabla^{\alpha *} = \nabla^{-\alpha}$ and this connection is torsion-free. Hence we have the same situation as in the classical case. Let $\alpha = \pm 1$ and let us consider the exponential family

$$\rho(\theta) = \exp\{\sum_{i=1}^m \theta_i A_i - \psi(\theta)\}.$$

Then it is $\pm 1$-flat and $\psi(\theta) = \log \operatorname{Tr} \exp(\theta_i A_i)$ is the potential function. Hence the coordinate systems $(\theta, \{\eta_i = \partial_i \psi(\theta)\}$ are mutually dual, $\theta$ is $\nabla^\alpha$-affine and $\eta$ is $\nabla^{-\alpha}$-affine. The divergence is given by

$$D_1(\rho_0, \rho_1) = \operatorname{Tr} \rho_1(\log \rho_1 - \log \rho_0)$$

which is the relative entropy. Similarly, for the mixture family

$$\rho(\eta) = \rho_0 + \sum_{i=1}^m \eta_i A_i, \qquad \operatorname{Tr} A_i = 0,$$

the function $\psi(\eta) = \operatorname{Tr} \rho(\eta) \log \rho(\eta)$ is the potential function, so that there is a pair of dual affine coordinate systems $(\eta, \theta)$, see also [9]. The divergence is

$$D_{-1}(\rho_0, \rho_1) = D_1(\rho_1, \rho_0).$$

For $\alpha \neq \pm 1$, we consider the extended $\alpha$-family

$$\rho(\theta) = g_\alpha^{-1}(\sum_{i+1}^m \theta_i A_i).$$

Let $\psi(\theta) = \frac{2}{1-\alpha} \operatorname{Tr} \rho(\theta)$. Then

$$\frac{\partial}{\partial \theta_i} \psi(\theta) = \frac{\partial}{\partial \theta_i} \frac{2}{1-\alpha} \operatorname{Tr} g_\alpha^{-1}(\sum_k \theta_k A_k) = \frac{2}{1-\alpha} \operatorname{Tr} L_\alpha^{-1} A_i = \operatorname{Tr} g_{-\alpha}(\rho(\theta)) A_i$$

and

$$\begin{aligned}\frac{\partial}{\partial \theta_j} \operatorname{Tr} g_{-\alpha}(\rho(\theta)) A_i &= \frac{\partial}{\partial \theta_j} \operatorname{Tr} g_{-\alpha}(g_\alpha^{-1}(\sum_k \theta_i A_k)) A_i \\ &= \operatorname{Tr} L_{-\alpha} L_\alpha^{-1}(A_j) A_k = \lambda^\alpha(A_i, A_j),\end{aligned}$$

hence $\psi(\theta)$ is the potential function. Thus there is a pair of dual affine coordinate systems and a divergence

$$D_\alpha(\rho_0, \rho_1) = \operatorname{Tr} g_\alpha(\rho_1)(g_{-\alpha}(\rho_1) - g_{-\alpha}(\rho_0)).$$

This $\alpha$-divergence was defined also in [8]. It is easy to see that the above divergence functions are the same as those from Proposition 6.2.

EXAMPLE 6.2. ([5, 13]) Let $\alpha = -1$ and let $\lambda$ be the metric of the symmetric logarithmic derivative. Then it is easy to see that

$$\rho_t = \exp\{\tfrac{1}{2}(tA - \psi(t))\} \rho_0 \exp\{\tfrac{1}{2}(tA - \psi(t))\},$$

where $\psi(t) = \log \mathrm{Tr}\, \rho_0 \exp(tA)$ is a solution of (i). Hence it is a $\nabla^{-1*}$-geodesic and the coordinate $t$ is $\nabla^{\alpha*}$-affine. It follows that $\rho_t$ is also $\nabla^{-1}$-flat and there is a $\nabla^{-1}$-affine coordinate system $s(t) = \mathrm{Tr}\, \rho_t A$. Further, the divergence is

$$D_{-1}(\rho_0, \rho_1) = 2\mathrm{Tr}\, \rho_1 \log \rho_0^{-\frac{1}{2}} (\rho_0^{\frac{1}{2}} \rho_1 \rho_0^{\frac{1}{2}})^{\frac{1}{2}} \rho_0^{-\frac{1}{2}}.$$

Clearly, this divergence coincides with the relative entropy if $\rho_0$ and $\rho_1$ commute. Moreover, $\rho_t$ has the Gibbs state $\exp(tA - \psi(t))$ as a special case.

EXAMPLE 6.3. Let $\alpha = -1$ and let $\lambda$ be the metric of the right logarithmic derivative. Then it is easy to see that

$$\tilde{\rho}_t = F^{\frac{1}{2}} \exp\{t Q_F(A)\} F^{\frac{1}{2}},$$

where $Q_F$ is a linear operator given by $Q_F^{-1}(A) = \frac{1}{2}(F^{-\frac{1}{2}} A F^{\frac{1}{2}} + F^{\frac{1}{2}} A F^{-\frac{1}{2}})$, is a solution of (10). If $\rho_0$ and $\rho_1$ are two states,

$$\rho_t = \rho_1^{\frac{1}{2}} \exp\{(t-1) Q_{\rho_1}(A) - \psi(t)\} \rho_1^{\frac{1}{2}},$$

with $\psi(t) = \log \mathrm{Tr}\, \rho_1 \exp\{(t-1) Q_{\rho_1}(A)\}$, is a $\nabla^{-1*}$-geodesic. Choose $A$ so that

$$\rho_1^{\frac{1}{2}} \exp\{-Q_{\rho_1}(A)\} \rho_1^{\frac{1}{2}} = \rho_0.$$

$\rho_t$ is then the $\nabla^{-1*}$-geodesic connecting these two states. We see that the divergence is given by

$$D^{-1}(\rho_0, \rho_1) = \mathrm{Tr}\, \rho_1 A = \mathrm{Tr}\, \rho_1 \log \rho_1^{\frac{1}{2}} \rho_0^{-1} \rho_1^{\frac{1}{2}}$$

This version of the relative entropy appeared also in [3].

## REFERENCES

[1] S. Amari: *Differential-geometrical methods in statistics*, Lecture Notes in Statistics **28**, Springer, Berlin 1985.

[2] S. Amari, O. E. Barndorff-Nielsen, R. E. Kass, S. L. Lauritzen, C. R. Rao: *Differential geometry in statistical inference*. IMS Lecture Notes—Monograph series 10, Mayward 1987.

[3] V. P. Belavkin and P. Staszewski: $C^*$-algebraic generalizations of relative entropy and entropy, *Ann. Inst. Henri Poincaré* **A37** (1982), 51–58.

[4] R. Bhatia: *Matrix analysis*, Springer, New York 1997.

[5] S. L. Braunstein and C. M. Caves: Geometry of quantum states. In: *Quantum Communication, Computing and Measurement* (Hirota et al., Eds.), Plenum Press, New York 1994.

[6] N. N. Chentsov: *Statistical decision rules and optimal inferences*, Translation of Math. Monog. **53**, Amer. Math. Society, Providence 1982.

[7] N. N. Chentsov and Morozova: Markov invariant geometry on state manifolds (in Russian), *Itogi Nauki i Tekhniki* **36** (1990), 69–102.

[8] H. Hasegawa: *Rep. Math. Phys.* **33** (1993), 87–93.

[9] H. Hasegawa: *Rep. Math. Phys.* **39** (1997), 49–68.

[10] H. Hasegawa and D. Petz: Non-commutative extension of information geometry II. In: *Quantum Communication, Computing and Measurement*, (eds. Hirota et al.), Plenum Press, New York 1997.

[11] A. S. Holevo: *Probabilistic and Statistical Aspects of Quantum Theory*, North-Holland, Amsterdam 1982.
[12] S. Kobayashi, K. Nomizu: *Foundation of Differential Geometry, Volume II*, Interscience, Wiley, New York, London 1963.
[13] H. Nagaoka: Differential geometrical aspects of quantum state estimation and relative entropy. In: *Quantum Communication, Computing and Measurement* (eds. Hirota et al.) Plenum Press, New York, 1994.
[14] D. Petz and Cs. Sudár: *J. Math. Phys.* **37** (1996), 2662–2673.
[15] D. Petz: Information geometry of quantum states. In: *Quantum Probability Communications*, vol. 10 (eds. R.L. Hudson et al,), World Scientific, 135–158, Singapore.
[16] D. Petz: *Linear Algebra Appl.* **244** (1996), 81–96.
[17] D. Petz and A. Jenčová: On quantum Fisher information, to appear.

International Journal of Theoretical Physics, Vol. 43, Nos. 7/8, August 2004 (© 2004)

# Generalized Relative Entropies as Contrast Functionals on Density Matrices

**Anna Jenčová**[1]

We use a class of generalized relative entropies on density matrices to obtain one-parameter families of torsion-free affine connections.

**KEY WORDS:** generalized relative entropies; information geometry; affine connections.

## 1. INTRODUCTION

The aim of quantum information geometry is to introduce the quantum counterparts of the basic structures of the classical theory, namely Riemannian metrics and families of affine connections. It is an important feature of the classical information manifolds, that if invariancy with respect to bijective transformations of the sample space is required, then these structures are unique (up to a multiplication factor): the Fisher metric and the family of Chentsov-Amari $\alpha$-connections (Amari, 1985; Chentsov, 1982).

Let $\mathcal{F} = \{p(\cdot, \theta) | \theta \in \Theta\}$ be a manifold of classical probability densities with respect to a common measure $P$. To define the affine connections, Amari (1985) used a family of functions

$$f_\alpha(x) = \begin{cases} \dfrac{2}{1-\alpha} x^{\frac{1-\alpha}{2}} & \alpha \neq 1 \\ \log(x) & \alpha = 1 \end{cases} \tag{1}$$

Let $l_\alpha(x, \theta) = f_\alpha(p(x, \theta))$. The coefficients of the Fisher information metric tensor and the $\alpha$-connections are given by

$$g_{ij}(\theta) = \int \partial_i l_\alpha(x, \theta) \partial_j l_{-\alpha}(x, \theta)\, dP, \quad \forall \alpha$$

[1] Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, SK-814 73 Bratislava, Slovakia; e-mail: jenca@mat.savba.sk.

$$\Gamma^{\alpha}_{ijk}(\theta) = \int \partial_i \partial_j l_\alpha(x,\theta) \partial_k l_{-\alpha}(x,\theta)\, dP$$

These connection are torsion-free and the $\alpha$ and $-\alpha$ connections are dual with respect to the Fisher metric, in the sense that if $\nabla^{\pm\alpha}$ are the covariant derivatives and $X, Y, Z$ are vector fields, then

$$Xg(Y,Z) = g\left(\nabla^{\alpha}_X Y, Z\right) + g\left(Y, \nabla^{-\alpha}_X Z\right)$$

There are more equivalent ways to introduce the Fisher metric and the affine connections. In the present paper, we follow the approach of Eguchi (1983), who used contrast functionals, see also Amari (1985).

A functional $\rho$ over $\mathcal{F} \times \mathcal{F}$ is called a contrast functional if

(i) $\rho(\theta_1, \theta_2) \geq 0$ for all $\theta_1, \theta_2 \in \Theta$
(ii) $\rho(\theta_1, \theta_2) = 0$ if and only if $\theta_1 = \theta_2$

The Riemannian metric and Christoffel symbols of the affine connections are defined by

$$g^{\rho}_{ij}(\theta) = -\frac{\partial^2}{\partial\theta_i \partial\theta'_j} \rho(\theta, \theta')|_{\theta=\theta'} \tag{2}$$

$$\Gamma^{\rho}_{ijk}(\theta) = -\frac{\partial^3}{\partial\theta_i \theta_j \theta'_k} \rho(\theta, \theta')|_{\theta=\theta'} \tag{3}$$

Let $f : (0, \infty) \to \mathbb{R}$ be a convex function satisfying $f(1) = 0$, then

$$\rho_f(\theta_1, \theta_2) = E_{\theta_1}\left[ f\left( \frac{p(X, \theta_2)}{p(X, \theta_1)} \right) \right]$$

defines a contrast functional. It was shown that in this case, $g^{\rho}_{ij} = f''(1) g_{ij}$, where $g_{ij}$ denotes the coefficients of the Fisher metric and the corresponding affine connection coincides with the $\alpha$-connection with $\alpha = 2f'''(1) + 3f''(1)$.

As one would expect, the situation is different in noncommutative case. Here, the equivalent of the Fisher metric would be a Riemannian metric, which is monotone with respect to completely positive trace preserving maps. For manifolds of $n \times n$ density matrices, it was proved by Chentsov and Morozova (1990) that such metric is not unique. Later, Petz (1996) characterized the class of monotone metrics in terms of operator monotone functions. Nagaoka (1994) defines the affine $\alpha$-connection for $\alpha = -1$ (the mixture connection) using the natural flat affine structure on density matrices. The exponential connection is defined as its dual with respect to the given monotone metric. This approach was generalized in Jenčová (2001a), for all $\alpha$. Unlike the classical case, the dual connections are not torsion free in general. In Jenčová (2001b), it was shown that the dual connection to the $\alpha$-connection is torsion-free only for a special monotone metric $\lambda^\alpha$.

Lesniewski and Ruskai (1999) used a class of generalized relative entropies, defined in Petz (1986), as contrast functionals on (non-normalized) density matrices. It was shown that each monotone metric can be obtained in the form (2) for a certain convex subset of relative entropies. The aim of the following paper is to use this subset to obtain a class of torsion free $\alpha$-connections, such that $\alpha$ and -$\alpha$-connections are dual. We question the coincidence with the Fisher metric and classical affine $\alpha$-connections on commutative submanifolds, use the language of statistical manifolds by Lauritzen (Amari *et al*., 1987), to give formulas for the Riemannian curvature tensor. We also treat some important examples.

## 2. GENERALIZED RELATIVE ENTROPIES AND MONOTONE METRICS

Let $\mathcal{D}$ denote the set of $n \times n$ complex Hermitian matrices and let $\mathcal{D}^+$ be the subset of positive definite matrices. As an open subset in $\mathcal{D}$, $\mathcal{D}^+$ inherits a natural affine parametrization and has the structure of a differentiable manifold. Let $T_\rho$ be the tangent space at $\rho$ and let $\lambda$ be the monotone Riemannian metric. Then $\lambda$ is of the form (Petz, 1996)

$$\lambda_\rho(X, Y) = \operatorname{Tr} X J_\rho(Y), \quad J_\rho^{-1} = f(L_\rho R_\rho^{-1}) R_\rho$$

where $f : (0, \infty) \to \mathbb{R}$ is an operator monotone function satisfying $f(t) = t f(t^{-1})$ and a normalization condition $f(1) = 1$, $L_\rho$ and $R_\rho$ are the left and right multiplication operator, respectively.

Let $\mathcal{G}$ be the set of operator convex functions $g : (0, \infty) \to \mathbb{R}$, satisfying $g(1) = 0$ and $g''(1) = 1$. It is known that each operator convex function with $g(1) = 0$ can be written in the form

$$g(w) = a(w-1) + b(w-1)^2 + c\frac{(w-1)^2}{w} + \int_0^\infty (w-1)^2 \frac{1+s}{w+s}\, d\mu(s) \quad (4)$$

where $b, c \geq 0$ and $\mu$ is a positive finite measure on $(0, \infty)$. The value of $a \in \mathbb{R}$ does not influence any of the following structures and therefore two functions in $\mathcal{G}$ that differ only in $a$ will be treated as equal.

Let $\mathcal{P}$ be the set of positive finite measures $\mu$ on $[0, \infty]$, such that $\int_{[0,\infty]} d\mu = \frac{1}{2}$. Then (4) establishes a one-to-one correspondence between $\mathcal{G}$ and $\mathcal{P}$, with $c = \mu(\{0\})$, $b = \mu(\{\infty\})$.

If $g$ is an operator convex function, we define its transpose $\hat{g}$ by $\hat{g}(w) = w g(w^{-1})$. It is clear that $\hat{g} \in \mathcal{G}$ if $g \in \mathcal{G}$ and that $g \mapsto \hat{g}$ induces the map $\mathcal{P} \to \mathcal{P}$, given by $\mu \mapsto \hat{\mu}$, where $d\hat{\mu}(s) = d\mu(s^{-1})$.

If $g = \hat{g}$, we say that $g$ is symmetric. The subset of symmetric functions in $\mathcal{G}$ will be denoted by $\mathcal{G}_{\text{sym}}$. Let $\sim$ be the equivalence relation on $\mathcal{G}$

$$g_1 \sim g_2 \iff g_1 + \hat{g}_1 = g_2 + \hat{g}_2.$$

The quotient space $\mathcal{G}|_\sim$ is isomorphic to $\mathcal{G}_{\text{sym}}$. Similarly, $\mathcal{P}_{\text{sym}}$ denotes the subset of measures symmetric with respect to the transform $s \mapsto s^{-1}$ and we have an equivalence relation $\sim$ on $\mathcal{P}$. Let us denote by $\mathcal{G}_h$ the equivalence class containing $\frac{1}{2}h$, where $\frac{1}{2}h \in \mathcal{G}_{\text{sym}}$, and similarly $\mathcal{P}_m$.

In Petz (1986), see also Lesniewski and Ruskai (1999), the following class of generalized relative entropies on $\mathcal{D}^+$ was introduced.

*Definition 2.1.* Let $g \in \mathcal{G}$. The relative $g$-entropy $H_g : \mathcal{D}^+ \times \mathcal{D}^+ \to \mathbb{R}$ is defined by

$$H_g(\rho, \sigma) = \mathrm{Tr}\, \rho^{\frac{1}{2}} g\left(\frac{L_\sigma}{R_\rho}\right)(\rho^{\frac{1}{2}})$$

**Proposition 2.1.** (Lesniewski and Ruskai, 1999).
*Let $g \in \mathcal{G}$ and let $a$, $b$, $c$ and $\mu$ be as above. Then*

$$\begin{aligned} H_g(\rho, \sigma) &= a\mathrm{Tr}\,(\sigma - \rho) \\ &\quad + \mathrm{Tr}\,(\sigma - \rho)\left\{b\rho^{-1} + c\sigma^{-1} + \int_0^\infty \frac{1+s}{L_\sigma + sR_\rho}\, d\mu(s)\right\}(\sigma - \rho) \\ &= a\mathrm{Tr}\,(\sigma - \rho) + \mathrm{Tr}\,(\sigma - \rho) R_\rho^{-1} k\left(\frac{L_\sigma}{R_\rho}\right)(\sigma - \rho) \end{aligned}$$

*where*

$$k(w) = \int_{[0,\infty]} \frac{1+s}{w+s}\, d\mu(s) = \frac{g(w) - a(w-1)}{(w-1)^2}.$$

The relative $g$-entropy can be used to define a Riemannian structure on $\mathcal{D}^+$ as follows. Let $X, Y \in T_\rho$, then

$$\lambda_\rho(X, Y) = -\frac{\partial^2}{\partial s \partial t} H_g(\rho + sX, \rho + tY)|_{s=t=0} = \mathrm{Tr}\, X R_\rho^{-1} k_{\text{sym}}\left(\frac{L_\rho}{R_\rho}\right)(Y)$$

where

$$k_{\text{sym}}(w) = k(w) + w^{-1}k(w^{-1}) = \frac{g(w) + \hat{g}(w)}{(w-1)^2}.$$

It was proved that this defines a monotone metric, where the corresponding operator monotone function is $f = 1/k$. Conversely, for a given monotone metric, we may put $g(w) = \frac{(w-1)^2}{f(w)}$. The condition $g''(1) = 1$ is equivalent to the normalization condition $f(1) = 1$. Thus we have

**Proposition 2.2.** *There is a one-to-one correspondence between monotone Riemannian metrics and equivalence classes $\mathcal{G}_h$.*

## 3. AFFINE CONNECTIONS

Let $\theta \in \Theta \subseteq \mathbb{R}^N$ be a smooth parameter in $\mathcal{D}^+$ and let $\partial_i = \frac{\partial}{\partial \theta_i}$. Let us fix a monotone Riemannian metric $\lambda$ on $\mathcal{D}^+$ and let $\mathcal{G}_h$ be the corresponding equivalence class. Let us choose a function $g \in \mathcal{G}_h$. In correspondence with the classical theory, we define the affine connections $\nabla^g$ by

$$\Gamma^g_{ijk}(\theta) = \lambda_\theta(\nabla_{\partial_i}\partial_j, \partial_k) = -\partial_i \partial_j \frac{\partial}{\partial \theta'_k} H_g(D(\theta), D(\theta'))|_{\theta=\theta'}$$

It is easy to show that this satisfies the transformation rules of an affine connection.

**Proposition 3.1.** *Let $g \in \mathcal{G}_h$. Then the connections $\nabla^g$ and $\nabla^{\hat{g}}$ are dual with respect to $\lambda$. Moreover, the connections are torsion-free.*

**Proof:** Consider the natural flat affine structure in $\mathcal{D}^+$ and let $X$ be a vector field, parallel with respect to this affine structure, then $X$ is constant over $\mathcal{D}^+$. As there is no danger of confusion, we will denote its value $X_\rho \in \mathcal{D}$ at $\rho$ by the same letter. Let $X, Y, Z$ be such vector fields. If $g \in \mathcal{G}_h$, then clearly $\hat{g} \in \mathcal{G}_h$ and $H_{\hat{g}}(\rho, \sigma) = H_g(\sigma, \rho)$, so that

$$\lambda_\rho\big(\nabla^{\hat{g}}_X Y, Z\big) = -\frac{\partial^3}{\partial t \partial s \partial u} H_g(\rho + uZ, \rho + sX + tY)|_{s=t=u=0}$$

Using the previous section, we get

$$\begin{aligned} X\lambda_\rho(Y, Z) &= \frac{d}{dt}\lambda_{\rho+tX}(Y, Z)|_{t=0} \\ &= \frac{d}{dt}\left(-\frac{\partial^2}{\partial s \partial u} H_g(\rho + tX + sY, \rho + tX + uZ)|_{s=u=0}\right)_{t=0} \\ &= \lambda_\rho\big(\nabla^g_X Y, Z\big) + \lambda_\rho\big(Y, \nabla^{\hat{g}}_X Z\big) \end{aligned}$$

so that the connections are dual. Torsion-freeness is obvious. □

Let $c_g : (0, \infty) \times (0, \infty) \to \mathbb{R}$ be given by $c_g(x, y) = \frac{1}{y} k(\frac{x}{y})$, where $k$ is as in Proposition 2.1. Note that $c_g(y, x) = c_{\hat{g}}(x, y)$ is obtained from $w^{-1}k(w^{-1})$ and that $c^{\mathrm{sym}}_g(x, y) = c_g(x, y) + c_g(y, x) = \frac{1}{y} k_{\mathrm{sym}}(\frac{x}{y})$ is the Morozova-Chentsov function. From Proposition 2.1, we get

$$H_g(\rho, \sigma) = a\mathrm{Tr}\,(\sigma - \rho) + \mathrm{Tr}\,(\sigma - \rho) c_g(L_\sigma, R_\rho)(\sigma - \rho) \tag{5}$$

For $\sigma, \rho \in \mathcal{D}^+$, the operator $c_g(L_\sigma, R_\rho)$ is positive on the space of $n \times n$ complex matrices with Hilbert-Schmidt inner product $\langle X, Y \rangle = \mathrm{Tr}\, X^* Y$.

**Proposition 3.2.** *Let*

$$T_s(X, Y, Z) = (1+s)\,\mathrm{Tr}\, X \frac{1}{sR_\rho + L_\rho}(Y)\frac{1}{R_\rho + sL_\rho}(Z)$$

*and let* $T_\infty = \lim_{s\to\infty} T_s$. *Then*

$$\lambda_\rho\big(\nabla^g_X Y, Z\big) = 2\int_{[0,\infty]} \Re T_s(Z, X, Y)\, d\mu(s) -$$

$$-2\int_{[0,\infty]} (\Re T_s(Y, X, Z) + \Re T_s(X, Y, Z))\, d\hat{\mu}(s)$$

**Proof:** From (5), we get

$$\begin{aligned}\lambda_\rho\left(\nabla^g_X Y, Z\right) = -\frac{d}{ds}\mathrm{Tr}\,\{&X c_g(L_{\rho+sZ}, R_\rho)(Y) + Y c_g(L_{\rho+sZ}, R_\rho)(X)\\ &- X c_g(L_\rho, R_{\rho+sY})(Z) - Z c_g(L_\rho, R_{\rho+sY})(X)\\ &- Y c_g(L_\rho, R_{\rho+sX})(Z) - Z c_g(L_\rho, R_{\rho+sX})(Y)\}|_{s=0}\end{aligned}$$

Further, for $\rho, \sigma \in \mathcal{D}^+$ and $X, Y \in \mathcal{D}$,

$$\begin{aligned}\mathrm{Tr}\, Y c_g(L_\sigma, R_\rho)(X) &= \langle Y, c_g(L_\sigma, R_\rho)(X)\rangle = \langle c_g(L_\sigma, R_\rho)(Y), X\rangle\\ &= \langle X, c_g(L_\sigma, R_\rho)(Y)\rangle^- = \mathrm{Tr}\, X c_g(L_\sigma, R_\rho)(Y)^-\end{aligned}$$

It follows that

$$\begin{aligned}\lambda_\rho\big(\nabla^g_X Y, Z\big) = -2\Re\mathrm{Tr}\,\Bigg\{&X\frac{d}{ds}c_g(L_{\rho+sZ}, R_\rho)(Y)\\ &-\left[X\frac{d}{ds}c_g(L_\rho, R_{\rho+sY})(Z) + Y\frac{d}{ds}c_g(L_\rho, R_{\rho+sX})(Z)\right]\Bigg\}\,|_{s=0}\end{aligned}$$

We have

$$c_g(x, y) = \mu(\{0\})x^{-1} + \mu(\{\infty\})y^{-1} + \int_0^\infty \frac{1+s}{x+sy}\, d\mu(s) \tag{6}$$

Let us first suppose that $\mu(\{0\}) = \mu(\{\infty\}) = 0$. Then we compute

$$\frac{d}{dt}c_g(L_{\rho+tZ}, R_\rho)|_0 = -\int_0^\infty (1+s)\frac{1}{L_\rho + sR_\rho}L_Z\frac{1}{L_\rho + sR_\rho}\, d\mu(s)$$

$$\frac{d}{dt}c_g(R_\rho, L_{\rho+tY})|_0 = -\int_0^\infty s(1+s)\frac{1}{R_\rho + sL_\rho}L_Y\frac{1}{R_\rho + sL_\rho}\, d\mu(s)$$

so that

$$-\frac{d}{dt}\mathrm{Tr}\, X c_g(L_{\rho+tZ}, R_\rho)(Y)|_0$$

$$= \int_0^\infty (1+s)\mathrm{Tr}\,\frac{1}{R_\rho + sL_\rho}(X)Z\frac{1}{sR_\rho + L_\rho}(Y)\,d\mu(s)$$

$$= \int_0^\infty T_s(Z, Y, X)\,d\mu(s)$$

and

$$-\frac{d}{dt}\mathrm{Tr}\,Xc_g(R_\rho, L_{\rho+tY})(Z)|_0$$

$$= \int_0^\infty s(1+s)\mathrm{Tr}\,\frac{1}{sR_\rho + L_\rho}(X)Y\frac{1}{R_\rho + sL_\rho}(Z)\,d\mu(s)$$

$$= \int_0^\infty (1+s)\mathrm{Tr}\,\frac{1}{R_\rho + sL_\rho}(X)Y\frac{1}{sR_\rho + L_\rho}(Z)\,d\hat{\mu}(s)$$

$$= \int_0^\infty T_s(Y, Z, X)\,d\hat{\mu}(s)$$

It follows that for each $s \in [0, \infty)$,

$$2\Re T_s(X, Y, Z) = T_s(X, Y, Z) + T_s(X, Z, Y)$$

so that $\Re T_s$ is a covariant 3-tensor, symmetric in last two variables. The statement now follows easily.

Let $\mu$ be concentrated in 0 and $\infty$. It is clear that $T_\infty = 0$ and we obtain by a direct computation from (6) that

$$\lambda_\rho\left(\nabla_X^g Y, Z\right) = \mu(\{0\})(T_0(Z, Y, X) + T_0(Z, X, Y)) - \mu(\{\infty\})(T_0(Y, X, Z) + T_0(Y, Z, X) + T_0(X, Y, Z) + T_0(X, Z, Y)) \qquad \square$$

### 3.1. Families of Connections

Let $\mathcal{G}_h$ be the equivalence class corresponding to the monotone metric $\lambda$. Let $g \in \mathcal{G}_h$. If $g$ is symmetric, then the connection $\nabla^g$ is self dual and torsion free, hence it is the metric connection. If $\lambda$ is fixed, we denote the metric connection by $\bar{\nabla}$.

Let $g \neq \hat{g}$. As $\mathcal{G}_h$ is a convex set, it contains all the functions

$$g_\alpha = \frac{1-\alpha}{2}g + \frac{1+\alpha}{2}\hat{g}$$

for $\alpha \in [-1, 1]$. If $\lambda$ and $g$ are fixed, we denote the corresponding connection by $\nabla^\alpha$. Then

$$\nabla^\alpha = \frac{1-\alpha}{2}\nabla + \frac{1+\alpha}{2}\nabla^*$$

where $\nabla$ and $\nabla^*$ are the covariant derivatives corresponding to $g$ and $\hat{g}$, respectively. The connections $\nabla^\alpha$ and $\nabla^{-\alpha}$ are dual with respect to $\lambda$, $\nabla^{-1} = \nabla$, $\nabla^1 = \nabla^*$ and $\nabla^0 = \bar{\nabla}$ for all $g$. Clearly, such family of $\alpha$-connections depends on the choice of $g \in \mathcal{G}_h$ and is therefore not unique.

### 3.2. Commutative Submanifolds

Let $\rho$, $X$, $Y$, and $Z$ be all mutually commuting. Then it is easy to see that $\lambda_\rho(X, Y) = \operatorname{Tr} \rho^{-1} XY$ and

$$\lambda_\rho\big(\nabla^g_X Y, Z\big) = -\frac{1+\alpha^*}{2} \operatorname{Tr} \rho^{-2} XYZ$$

where $\alpha^* = 2g'''(1) + 3$. This corresponds to the Fisher metric and the $\alpha^*$-connection in the commutative case. It seems to be a natural question to ask if, for each $\lambda$, it is possible to obtain the $\alpha^*$-connections at least for $\alpha^* \in [-1, 1]$, if restricted to commutative submanifolds. From the next proposition (and examples below) it follows that this is not true.

Let the Riemannian metric $\lambda$ correspond to the equivalence class $\mathcal{G}_h$, resp. $\mathcal{P}_m$. Let $\mu_{\max}$ be a measure with supp $\mu \subseteq [1, \infty]$, such that $\mu_{\max}$ coincides with $m$ on $(1, \infty]$ and $\mu_{\max}(\{1\}) = \frac{1}{2} m(\{1\})$. Then we have

**Proposition 3.3.** *Let $\mu_{\max}$ be as above and let $g_{\max}$ be the corresponding operator convex function. Then $g_{\max} \in \mathcal{G}_h$ and for each $g \in \mathcal{G}_h$, we have*

$$-3 \leq \hat{g}'''_{\max}(1) \leq g'''(1) \leq g'''_{\max}(1) \leq 0$$

**Proof:** First, it is easy to see that $\mu_{\max}$ is a positive finite measure and $\int_{[0,\infty]} d\mu_{\max} = \frac{1}{2} \int_{[0,\infty]} dm = \frac{1}{2}$. Moreover, $\hat{\mu}_{\max}$ is concentrated in $[0, 1]$, $\hat{\mu}_{\max}$ coincides with $m$ on $[0, 1)$ and $\hat{\mu}_{\max}(\{1\}) = \frac{1}{2} m(\{1\})$, so that $\mu_{\max} + \hat{\mu}_{\max} = m$. It follows that $g_{\max} \in \mathcal{G}_h$. Let now $g \in \mathcal{G}_h$ and let $\mu \in \mathcal{P}_m$ be the corresponding measure. Then

$$g'''(1) = -6 \int_{[0,\infty]} \frac{1}{1+s}\, d\mu(s)$$

and

$$\begin{aligned}
\int_{[0,\infty]} \frac{1}{1+s}\, d\mu(s) &= \int_{(1,\infty]} \frac{s}{1+s}\, d\mu(s^{-1}) + \frac{1}{2}\mu(\{1\}) + \int_{(1,\infty]} \frac{1}{1+s}\, d\mu(s) \\
&\geq \int_{(1,\infty]} \frac{1}{1+s}\, dm(s) + \frac{1}{4} m(\{1\}) = \int_{[0,\infty]} \frac{1}{1+s}\, d\mu_{\max} \geq 0
\end{aligned}$$

and similarly,

$$\int_{[0,\infty]} \frac{1}{1+s} d\mu(s) \leq \int_{[0,\infty]} \frac{1}{1+s} d\hat{\mu}_{\max}(s) \leq \frac{1}{2} \qquad \square$$

## 4. EXAMPLE 1: THE EXTREME BOUNDARY OF $\mathcal{G}$

The extreme boundary of $\mathcal{G}$ consists of the functions

$$g_s(w) = \frac{1+s}{2}\frac{(w-1)^2}{w+s} \text{ for } s \geq 0$$

$$g_\infty(w) = \frac{1}{2}(w-1)^2$$

We have $\hat{g}_s = g_{s^{-1}}$ for $s > 0$ and $\hat{g}_0 = g_\infty$, $g_1$ being the only symmetric one of these functions. The corresponding measures are $\mu_s(t) = \frac{1}{2}\delta(s-t)$.

Let $s \in [0, 1]$. Denote $h_s = g_s + \hat{g}_s$, then

$$h_s(w) = \frac{(1+s)^2}{2}(w-1)^2 \frac{w+1}{(w+s)(sw+1)}$$

Let $\lambda_s$ be the corresponding monotone metric. It is easy to see that $g_{s,max} = \hat{g}_s$ and that

$$\mathcal{G}_s := \mathcal{G}_{h_s} = \left\{ g_\alpha = \frac{1-\alpha}{2} g_s + \frac{1+\alpha}{2} g_{s^{-1}} : \alpha \in [-1, 1] \right\}$$

In particular, $\mathcal{G}_1 = \{g_1\}$. It follows that for each $\lambda_s$, we have a unique family of $\alpha$-connections. If we consider commutative submanifolds, we obtain classical $\alpha^*$-connections with $\alpha^* \in [-3\frac{1-s}{1+s}, 3\frac{1-s}{1+s}]$. Two important special cases, $s = 1$ and $s = 0$ will be treated below.

### 4.1. The Metric of Bures

Let us consider the previous example with $s = 1$. Then

$$h_1(w) = 2\frac{(w-1)^2}{w+1}$$

and the corresponding monotone metric is given by

$$\lambda_{1\rho}(X, Y) = \mathrm{Tr}\, X \frac{2}{L_\rho + R_\rho}(Y)$$

It is the smallest metric in the class of monotone metrics. We have already seen that the corresponding equivalence class consists of only one function $g_1$. It means that the only connection that we can obtain is the metric connection $\bar{\nabla}$.

### 4.2. The Largest Monotone Metric

Let $s = 0$. Then

$$h_0(w) = \frac{1}{2}(w-1)^2\frac{w+1}{w}$$

and $\lambda_0$ is given by

$$\lambda_{0\rho}(X, Y) = \mathrm{Tr}\, X\frac{1}{2}\left(R_\rho^{-1} + L_\rho^{-1}\right)(Y)$$

It is the largest monotone metric. On commutative submanifolds, we obtain $\alpha^*$ -connections for $\alpha^*$ in the largest possible interval $\alpha^* \in [-3, 3]$. It is easy to see from Proposition 3.3 that this is the only monotone metric with this property.

## 5. STATISTICAL MANIFOLDS

The manifold $\mathcal{D}^+$ with a monotone metric and a class of $\alpha$-connections can be regarded as a statistical manifold in the sense of Lauritzen (Amari *et al.*, 1987). A statistical manifold is a triple $(M, g, \tilde{D})$, where $M$ is a differentiable manifold, $g$ a metric tensor and $\tilde{D}$ a symmetric covariant 3-tensor, called the skewness of the manifold. On $M$, a class of $\alpha$-connections is introduced by

$$\nabla_X^\alpha Y = \bar{\nabla}_X Y - \frac{\alpha}{2}D(X, Y), \tag{7}$$

where $\bar{\nabla}$ is the metric connection and the tensor $D$ is defined by $\tilde{D}(X, Y, Z) = g(D(X, Y), Z)$. These connections are torsion free, this is equivalent to symmetry of $\tilde{D}$, resp. $D$. The Riemannian curvature tensor is defined as

$$R^\alpha(X, Y, Z, W) = g\left(\nabla_X^\alpha\nabla_Y^\alpha Z - \nabla_Y^\alpha\nabla_X^\alpha Z - \nabla_{[X,Y]}^\alpha Z, W\right)$$

Statistical manifolds satisfying $R^\alpha = R^{-\alpha}$ for all $\alpha$ are called conjugate symmetric. It is proved that $R^{-\alpha} - R^\alpha = \alpha\{F(X, Y, Z, W) - F(Y, X, Z, W)\}$, where $F(X, Y, Z, W) = (\bar{\nabla}_X\tilde{D})(Y, Z, W)$, so that a statistical manifold is conjugate symmetric if and only if the tensor $F$ is symmetric. It also follows that the condition

$$\exists\alpha_0 \neq 0, \quad R^{\alpha_0} = R^{-\alpha_0}$$

is sufficient for conjugate symmetry.

Let now $\lambda$ be a monotone metric on $\mathcal{D}^+$ and let $\mathcal{G}_h$ be the corresponding equivalence class. Let $g \in \mathcal{G}_h$ such that $g$ is not symmetric and let us consider the corresponding family of connections. Let

$$D(X, Y) = \nabla_X Y - \nabla_X^* Y$$

Then the triple $(\mathcal{D}^+, \lambda, \tilde{D})$ is a statistical manifold, with $\tilde{D}(X, Y, Z) = \lambda(D(X, Y), Z)$, and the family of connections has the form (7).

Let $K$ be a covariant $k$-tensor field, then its symmetrization $K^{\text{sym}}$ is defined as

$$K^{\text{sym}}(X_1, \ldots, X_k) = \frac{1}{k!} \sum_{\pi} K\left(X_{\pi(1)}, \ldots, X_{\pi(k)}\right)$$

where $\pi$ runs over all permutations of the set $\{1, \ldots, k\}$.

**Proposition 5.1.** *Let $T_s^{\text{sym}}$ be the symmetrization of $\Re T_s$. Then $\tilde{D}$ has the form*

$$\tilde{D}(X, Y, Z) = 6 \int_{[0,\infty]} T_s^{\text{sym}}(X, Y, Z)\, d(\mu - \hat{\mu})(s)$$

**Proof:** Straightforward from Proposition 3.2. □

Let us now compute the Riemannian curvature tensor $R^{\alpha}$ of the $\alpha$-connection.

**Proposition 5.2.** *Let $X, Y, Z, W$ be vector fields on $\mathcal{M}^+$ and let $\bar{R} = R^0$. Then*

$$\begin{aligned} R^{\alpha}(X, Y, Z, W) = \bar{R}(X, Y, Z, W) &+ \frac{\alpha}{2}\{F(Y, X, Z, W) - F(X, Y, Z, W)\} \\ &+ \frac{\alpha^2}{4}\{\lambda(D(X, W), D(Y, Z)) - \lambda(D(X, Z), D(Y, W))\} \end{aligned}$$

**Proof:** As we are going to establish a tensorial equality, we may suppose that $[X, Y] = 0$. As the metric connection is symmetric, we have $\bar{\nabla}_X Y - \bar{\nabla}_Y X = 0$. From (7) we obtain, using symmetry of $\tilde{D}$

$$\begin{aligned} \lambda\left(\nabla_X^{\alpha} \nabla_Y^{\alpha} Z, W\right) = \lambda(\bar{\nabla}_X \bar{\nabla}_Y Z, W) &- \frac{\alpha}{2}\{\lambda(\bar{\nabla}_X D(Y, Z), W) + \lambda(D(X, \bar{\nabla}_Y Z), W)\} \\ &+ \frac{\alpha^2}{4}\lambda(D(X, W), D(Y, Z)) \end{aligned}$$

Subtracting the expression with interchanged $X$ and $Y$ and using self-duality and symmetry of $\bar{\nabla}$ completes the proof. □

**Corollary 5.1.** *Let the manifold be conjugate symmetric. Then we have*

$$\begin{aligned} R^{\alpha}(X, Y, Z, W) &= \bar{R}(X, Y, Z, W) \\ &\quad + \frac{\alpha^2}{4}\{\lambda(D(X, W), D(Y, Z)) - \lambda(D(X, Z), D(Y, W))\} \end{aligned}$$

*If $\theta \mapsto \rho(\theta)$ is a smooth parametrization of $\mathcal{D}^+$, then*

$$R^{\alpha}_{ijkl}(\theta) = \bar{R}_{ijkl}(\theta) + \frac{\alpha^2}{4} \sum_{\beta,\gamma} (\tilde{D}_{il\beta}\tilde{D}_{jk\gamma} - \tilde{D}_{ik\beta}\tilde{D}_{jl\gamma})\lambda^{\beta\gamma}$$

*where $\lambda^{ij} = (\lambda^{-1})_{ij}$.*

**Corollary 5.2.** *If* $\exists\alpha_0 \neq 0$*, such that* $R^{\alpha_0} = 0$*, then*

$$R^\alpha(X, Y, Z, W) = \frac{\alpha^2 - \alpha_0^2}{4}\{\lambda(D(X, W), D(Y, Z)) - \lambda(D(X, Z), D(Y, W))\}$$

*for* $\forall\alpha$*. Moreover, there exists a parametrization,* $\theta \mapsto \rho(\theta)$*, such that*

$$R^\alpha_{ijkl} = \frac{\alpha^2 - \alpha_0^2}{4\alpha_0^2}\sum_{\beta,\gamma}(\Gamma_{il\beta}\Gamma_{jk\gamma} - \Gamma_{ik\beta}\Gamma_{jl\gamma})\lambda^{\beta\gamma}$$

*where* $\Gamma_{ijk} = \lambda(\nabla^{-\alpha_0}_{\partial_i}\partial_j, \partial_k)$ *are the Christoffel symbols of* $\nabla^{-\alpha_0}$.

**Proof:** The connections $\nabla^\alpha$ and $\nabla^{-\alpha}$ are mutually dual, therefore $0 = R^{\alpha_0} = R^{-\alpha_0}$. It follows that the manifold is conjugate symmetric and we may use Corollary 5.1.

Further, let us define

$$D_{\alpha_0}(X, Y) = \nabla^{-\alpha_0} - \nabla^{\alpha_0},$$

then $D_{\alpha_0} = \alpha_0 D$ and

$$\nabla^\alpha = \bar{\nabla} - \frac{\alpha}{2\alpha_0}D_{\alpha_0}$$

It follows that

$$R^\alpha(X, Y, Z, W) = \frac{\alpha^2 - \alpha_0^2}{4\alpha_0^2}\{\lambda(D_{\alpha_0}(X, W), D_{\alpha_0}(Y, Z)) - \lambda(D_{\alpha_0}(X, Z), D_{\alpha_0}(Y, W))\}$$

As the manifold is $\pm\alpha_0$-flat, there exists an $\alpha_0$-affine parametrization $\theta \mapsto \rho(\theta)$, i.e. such that $\nabla^{\alpha_0}_{\partial_i}\partial_j = 0$ for all $i, j$. It follows that

$$\tilde{D}^{\alpha_0}_{ijk} = \lambda(D_{\alpha_0}(\partial_i, \partial_j), \partial_k) = \Gamma_{ijk}, \quad \forall i, j, k$$ □

**Corollary 5.3.** *If* $\exists\alpha_1 \neq \pm\alpha_2$ *such that* $R^{\alpha_1} = R^{\alpha_2} = 0$*, then* $R^\alpha = 0$ *for all* $\alpha$.

**Proof:** We may suppose that $\alpha_1 \neq 0$ and use Corollary 5.2. □

## 6. EXAMPLE 2: $\alpha$-DIVERGENCES

Let

$$g_\alpha = \begin{cases} \frac{4}{1-\alpha^2}\left(\frac{1+w}{2} - w^{\frac{1+\alpha}{2}}\right) & \alpha \neq \pm 1 \\ -\log w & \alpha = -1 \\ w \log w & \alpha = 1 \end{cases}$$

Then $g_\alpha \in \mathcal{G}$ for $\alpha \in [-3, 3]$. Moreover, $\hat{g}_\alpha = g_{-\alpha}$. The corresponding family of relative entropies and monotone metrics was defined by Hasegawa (1993). We have

$$\lambda_\alpha(X, Y) = \frac{\partial^2}{\partial s \partial t} \operatorname{Tr} f_\alpha(\rho + sX) f_{-\alpha}(\rho + tY)|_{t=s=0}$$

where $f_\alpha$ is the family of functions defined in Section 1. It is easy to show that the corresponding affine connections $\nabla^{g_\alpha}$ coincide with the $\alpha$-connections for $\lambda_\alpha$ defined in Jenčová (2001a). As the connections are torsion-free, this is the only case when this may happen, see also Jenčová (2001b).

There are some important special cases. For $\alpha = \pm 1$ we obtain the well known Bogoljubov-Kubo-Mori (BKM) metric. Another important example is $\alpha = \pm 3$, corresponding to the largest monotone metric, see Example 4.2. This is the unique monotone metric that is contained in both classes $\lambda_\alpha$ and $\lambda_s$ from Section 4.

Let us fix $\alpha_0 \in (0, 3]$. Then

$$h_{\alpha_0}(w) = g_{\alpha_0}(w) + g_{-\alpha_0}(w) = \frac{4}{1 - \alpha_0^2}\left(1 - w^{\frac{1-\alpha_0}{2}}\right)\left(1 - w^{\frac{1+\alpha_0}{2}}\right)$$

If we proceed as in the proof of Corrolary 5.2, we see that the family of connections

$$\nabla^\alpha = \bar{\nabla} - \frac{\alpha}{2\alpha_0} D_{\alpha_0},$$

can be obtained from $\mathcal{G}_{\alpha_0} = \mathcal{G}_{h_{\alpha_0}}$ for $\alpha \in [-\alpha_0, \alpha_0]$. In particular, $\nabla^{\alpha_0} = \nabla^{g_{\alpha_0}}$. As it was shown in Jenčová (2001a), the connection $\nabla^{\pm\alpha_0}$ is flat, i.e. the Riemannian curvature tensor $R^{\pm\alpha_0}$ vanishes. Hence, for the -$\alpha_0$-affine parametrization $\theta$,

$$\begin{aligned} R^\alpha_{ijkl} &= \frac{\alpha^2 - \alpha_0^2}{4\alpha_0^2}\left\{\lambda_{\alpha_0}\left(\nabla^{\alpha_0}_{\partial_i}\partial_l, \nabla^{\alpha_0}_{\partial_j}\partial_k\right) - \lambda_{\alpha_0}\left(\nabla^{\alpha_0}_{\partial_i}\partial_k, \nabla^{\alpha_0}_{\partial_j}\partial_l\right)\right\} \\ &= \frac{\alpha^2 - \alpha_0^2}{4\alpha_0^2}\sum_{\beta,\gamma}\left(\Gamma^{\alpha_0}_{il\beta}\Gamma^{\alpha_0}_{jk\gamma} - \Gamma^{\alpha_0}_{ik\beta}\Gamma^{\alpha_0}_{jl\gamma}\right)\lambda^{\beta\gamma} \end{aligned}$$

where

$$\Gamma^{\alpha_0}_{ijk} = \operatorname{Tr} \partial_i \partial_j f_{\alpha_0}(\rho) \partial_k f_{-\alpha_0}(\rho)$$

In particular, for $\alpha_0 = 1$ (the BKM metric), $\nabla^{g_{-1}}$ and $\nabla^{g_1}$ correspond to the mixture and exponential connections $\nabla^{(m)}$ and $\nabla^{(e)}$, respectively. The $\alpha$-connection is then a convex mixture of the $(m)$ and $(e)$-connections,

$$\nabla^\alpha = \frac{1-\alpha}{2}\nabla^{(m)} + \frac{1+\alpha}{2}\nabla^{(e)}$$

In the commutative case, this is an equivalent definition of the $\alpha$-connections. If we consider the natural affine parametrization $\rho(\theta) = \rho_0 + \sum_i \theta_i X_i$, the coefficients of Riemannian curvature tensor can be written in the form

$$R^{\alpha}_{ijkl} = \frac{\alpha^2 - 1}{4} \mathrm{Tr} \int_0^1 \{\partial_i \partial_j \log(\rho) \rho^t \partial_j \partial_k \log(\rho) \rho^{1-t} - \partial_i \partial_k \log(\rho) \rho^t \ \partial_j \partial_l \log(\rho) \rho^{1-t}\} \, dt$$

If $\{X_i\}$ is an orthonormal basis of $\mathcal{D}$ with respect to the metric $\lambda^{BKM}_{\rho_0}$, we may compute the coefficients at $\theta = 0$ as

$$R^{\alpha}_{ijkl} = \frac{\alpha^2 - 1}{4} \sum_{\beta} (\Gamma_{il\beta} \Gamma_{jk\beta} - \Gamma_{ik\beta} \Gamma_{jl\beta})$$

where

$$\Gamma_{ijk} = \mathrm{Tr}\, \partial_i \partial_j \log(\rho) X_k = -\mathrm{Tr}\, X_k \int_0^{\infty} [(\rho_0 + s)^{-1} X_i (\rho_0 + s)^{-1} X_j (\rho_0 + s)^{-1} + (\rho_0 + s)^{-1} X_j (\rho_0 + s)^{-1} X_i (\rho_0 + s)^{-1}] \, ds$$

As it was already proved e.g. in Petz (1994), the Riemannian curvature $\bar{R}$ of the metric connection given by $\lambda^{BKM}$ is not equal to 0. Using Corrolary 5.3, it follows that $R^{\alpha} = 0$ if and only if $\alpha = \pm 1$.

## ACKNOWLEDGMENT

Research supported by the grant VEGA 1/0264/03.

## REFERENCES

Amari, S. (1985). Differential-geometrical methods in statistic, *Lecture Notes in Statistics*, Vol. 28.

Amari, S., Barndorff-Nielsen, O. E., Kass, R. E., Lauritzen, S. L., and Rao, C. R. (1987). Differential geometry in statistical inference, *IMS Lecture notes-Monograph series*, Vol. 10.

Chentsov, N. N. (1982). Statistical decision rules and optimal inferences, *Translation of Mathematical Monograph 53*, American Mathematical Society, Providence.

Chentsov, N. N. and Morozova, E. A. (1990). Markov invariant geometry on state manifolds, *Itogi Nauki i Tekhniki* **36**, 69–102, (in Russian).

Eguchi, S. (1983). Second order efficiency of minimum contrast estimation in a curved exponential family, *Annals of Statistics* **11**, 793–803.

Hasegawa, H. (1993). $\alpha$-divergence of the non-commutative information geometry, *Reports on Mathematical Physics* **33**, 87–93.

Jenčová, A. (2001). Geometry of quantum states: dual Connections and divergence functions. *Reports on Mathematical Physics* **47**, 121–138.

Jenčová, A. (2001). Dualistic properties of the manifold of quantum states. In *Disordered and Complex Systems*, P. Sollich *et al.*, eds., AIP, Melville, New York.

Lesniewski, L. and Ruskai, M. B. (1999). Monotone Riemannian metrics and relative entropy on noncommutative probability spaces, *Journal of Mathematical Physics* **40**, 5702–5724.

Nagaoka, H. (1994). Differential geometrical aspects of quantum state estimation and relative entropy. In *Quantum Communication, Computing and Measurement*, Hirota *et al.*, eds., Plenum Press, New York.

Petz, D. (1986). Quasi-entropies for finite quantum systems. *Reports on Mathematical Physics* **23**, 57–65.

Petz, D. (1994). Geometry of canonical correlation on the state space of a quantum system. *Journal of Mathematical Physics* **244**, 780–795.

Petz, D. (1996). Monotone metrics on matrix spaces, *Linear Algebra and its Applications* **244**, 81–96.

# FLAT CONNECTIONS AND WIGNER–YANASE–DYSON METRICS

ANNA JENČOVÁ

Mathematical Institute, Slovak Academy of Sciences
Štefánikova 49, SK-814 73 Bratislava, Slovakia
(e-mail: jenca@mat.savba.sk)

(*Received October — Revised July 29, 2003*)

On the manifold of positive definite matrices, we investigate the existence of pairs of flat affine connections, dual with respect to a given monotone metric. The connections are defined either using the $\alpha$-embeddings and finding the duals with respect to the metric, or by means of contrast functionals. We show that in both cases, the existence of such a pair of connections is possible if and only if the metric is given by the Wigner–Yanase–Dyson skew information.

**Keywords:** monotone metrics, flat affine connections, duality, generalized relative entropies, WYD metrics.

## 1. Introduction

An important feature of the classical information geometry is the uniqueness of its structures, the Fisher metric and the family of affine $\alpha$-connections on a manifold $\mathcal{P}$ of probability distributions [5, 1]. In the case of finite quantum systems, this uniqueness does not take place: it was shown by Chentsov and Morozova [6] and later by Petz [22] that there are infinitely many Riemannian metrics, which are monotone with respect to stochastic maps. As for the affine connections, there were several definitions of the $\alpha$-connections [16, 19, 12, 14].

In the commutative case, two equivalent definitions of the connections were used by Amari [1]. First, the connections can be defined using $\alpha$-embeddings ($\alpha$-representations) given by the family of functions

$$f_\alpha(x) = \begin{cases} \dfrac{2}{1-\alpha} x^{\frac{1-\alpha}{2}}, & \alpha \neq 1, \\ \log(x), & \alpha = 1. \end{cases} \tag{1}$$

On the other hand, the connections can be defined as mixtures of the exponential and the mixture connections,

$$\nabla^{(\alpha)} = \frac{1+\alpha}{2} \nabla^{(e)} + \frac{1-\alpha}{2} \nabla^{(m)}. \tag{2}$$

Such connections are torsion-free and the $\alpha$ and $-\alpha$ connections are dual with respect to the Fisher metric. Moreover, in the case of a finite system, that is on the manifold of all (non-normalized) multinomial distributions, the $\alpha$-connections are flat for all $\alpha$.

The definition involving $\alpha$-representations can be easily generalized to noncommutative case to obtain a family of flat connections $\nabla^{(\alpha)}$ on the manifold of positive definite matrices. This definition was treated also by the present author in [16] and [17]. The dual of such $\alpha$-connection with respect to a given monotone metric is in general different from the $-\alpha$-connection. The duals have vanishing Riemannian curvature, but are not always torsion-free and hence not flat. The condition that the dual of the $\alpha$-connection with respect to a monotone metric is torsion-free restricts $\alpha$ to the interval $[-3, 3]$ and, for such $\alpha$, singles out a monotone metric $\lambda_\alpha$, which belongs to the family of Wigner–Yanase–Dyson (WYD) metrics. This is also equivalent to the condition that the dual of $\nabla^{(\alpha)}$ is $\nabla^{(-\alpha)}$, see also [10]. A brief description of these results is given in Sections 2 and 3.

For $\alpha = \pm 1$, we get the Kubo–Mori–Bogoljubov metric, with respect to which the mixture $\nabla^{(m)}$ and exponential $\nabla^{(e)}$ connections are dual. As in the classical case, we may use mixtures of $\nabla^{(e)}$ and $\nabla^{(m)}$ to define a family of torsion-free connections, having the required duality properties with respect to the BKM metric. In our approach, however, the value of $\alpha$ in (2) will be restricted to the interval $[-1, 1]$, but the proofs in Section 5 suggest that our results hold more generally. Convex mixtures were considered also by Grasselli and Streater, see the Discussion in [11]. We will answer the questions discussed there in proving that, for $\alpha \in (-1, 1)$, affine connections defined by (2) are different from the $\alpha$-connections and are not flat. A simple direct proof of this fact can be found at the end of Section 5.

Another way to define an affine connection was proposed by Eguchi in [9], by means of a contrast functional on $\mathcal{P}$. A functional $\phi : \mathcal{P} \times \mathcal{P} \to \mathbb{R}$ is a contrast functional if it satisfies $\phi(p, q) \geq 0$ for all $p, q$ and $\phi(p, q) = 0$ if and only if $p = q$. Using such a functional, a metric tensor and affine connection can be defined. Let $\theta_1, \dots, \theta_p$ be a smooth parametrization of $\mathcal{P}$ and let $\partial_i$, $i = 1, \dots, p$, be the corresponding vector fields, then the metric tensor is given by

$$g_{ij}^{\phi} = -\partial_i \partial_j' \phi(p(\theta), p(\theta'))|_{\theta=\theta'}.$$

An affine connection $\nabla^{\phi}$ is defined by

$$\Gamma_{ijk}^{\phi} = g^{\phi}(\nabla_{\partial_i}^{\phi} \partial_j, \partial_k) = -\partial_i \partial_j \partial_k' \phi(p(\theta), p(\theta'))|_{\theta=\theta'}.$$

Consider a special class of contrast functionals $\phi_g$, related to convex functions $g$ satisfying $g(1) = 0$ by

$$\phi_g(p, q) = \int g\left(\frac{q}{p}\right) dp.$$

In this case, it was shown [1] that the corresponding metric is the Fisher metric (multiplied by $g''(1)$) and the affine connection is the $\alpha$-connection, $\alpha = 2g'''(1) + 3$.

As the quantum counterpart of such functionals we will use the relative $g$-entropies $H_g$ defined by Petz [20],

$$H_g(\rho, \sigma) = \operatorname{Tr} \rho^{1/2} g(L_\sigma / R_\rho)(\rho^{1/2}),$$

where $g$ is an operator convex function and $g(1) = 0$. It was shown that [18]:

(a) In the normalized case (or if $\operatorname{Tr} \rho = \operatorname{Tr} \sigma$), $H_g(\rho, \sigma) \geq 0$ and $H_g(\rho, \sigma) = 0$ if and only if $\rho = \sigma$.
(b) $H_g(\lambda\rho, \lambda\sigma) = \lambda H_g(\rho, \sigma)$ for each $\lambda > 0$.
(c) $H_g$ is jointly convex in $\rho$ and $\sigma$.
(d) $H_g$ is monotone, that is, it decreases under stochastic maps.
(e) $H_g$ is differentiable.

We see that $H_g$ is a contrast functional on the manifold of quantum states, and we will show that we can use it to define the geometrical structures as above, even in the non-normalized case. The relative $g$-entropies were used by Lesniewski and Ruskai [18], who proved that the Riemannian structure given by $H_g$ is monotone for each $g$ and, conversely, each monotone metric is obtained in this way. A short account on some of their results is in Section 4.

In Section 5 we will use $H_g$ to define an affine connection and show that this definition contains both the $\alpha$-connections, defined from $\alpha$-embeddings, and the convex mixtures of $\nabla^{(m)}$ and $\nabla^{(e)}$. We will show that for each monotone metric there is a family of such connections (the $p$-connections) parametrized by $p \in [0, 1]$, such that they are torsion-free and the $p$- and $(1-p)$-connections are dual. We will then use the theory of statistical manifolds by Lauritzen [2] to investigate the Riemannian curvature of the connections.

Finally, in the last section we will show that a pair of dual flat connections exists if and only if the metric is one of the WYD metrics $\lambda_\alpha$. The flat connections are then the $\pm\alpha$-connections. This result holds for the connections given by the relative $g$-entropies. It is known from [1] that dual flat connections give rise to divergence functionals on the manifold, it is therefore reasonable to consider connections defined from functionals having the properties (a)–(e). The class of $g$-entropies seems to be large enough, although it does not contain all such functionals (see [18]). The main results of the present paper can be summarized as follows: If a pair of dual flat connections is required, the structures of information geometry are unique even in the quantum case, at least if we consider only connections defined by the relative $g$-entropies. These structures are provided by the family of Wigner–Yanase–Dyson metrics and the $\alpha$-connections.

## 2. The manifold and monotone metrics

Let $\mathcal{M}_n(\mathbb{C})$ be the space of $n \times n$ complex matrices, $\mathcal{M}_h$ be the real linear subspace of hermitian matrices and let $\mathcal{M} \subset \mathcal{M}_h$ denote the set of positive definite matrices. As an open subset in a finite-dimensional real vector space, $\mathcal{M}$ inherits the structure of a differentiable manifold. The tangent space $T_\rho$ of $\mathcal{M}$ at $\rho$ is the linear space of directional (Frèchet) derivatives in the direction of smooth curves in $\mathcal{M}$ and it can be identified with $\mathcal{M}_h$ in an obvious way. In the present paper, the elements of the tangent space, seen as directional derivative operators, will be denoted by $\mathcal{X}, \mathcal{Y}$, etc., while the corresponding capital letters will mean their representations $X = \mathcal{X}(\rho)$ etc. in $\mathcal{M}_h$. The map $\mathcal{X} \mapsto X$ is the same as Amari's -1-representation of the tangent space in the classical case [1], see also the next section. The vector fields on $\mathcal{M}$ are represented by $\mathcal{M}_h$-valued functions on $\mathcal{M}$. If $\mathcal{X}, \mathcal{Y}$ are vector fields, then the bracket $[\mathcal{X}, \mathcal{Y}]$ is unrelated to the usual commutator of the representing matrices and these two should not be confused. In the present paper we will use $[\cdot, \cdot]$ only in the first (vector fields) meaning.

A Riemannian structure is introduced in $\mathcal{M}$ by

$$\lambda_\rho(X, Y) = \operatorname{Tr} X J_\rho(Y), \qquad X, Y \in T_\rho,$$

where $J_\rho$ is a suitable operator on matrices. We say that the metric $\lambda$ is monotone if it is monotone with respect to stochastic maps, that is, we have

$$\lambda_{T(\rho)}(T(X), T(X)) \leq \lambda_\rho(X, X), \qquad \rho \in \mathcal{M}, \ \ X \in T_\rho,$$

for a stochastic map $T$. It is an important result of Petz [22], that this is equivalent to

$$J_\rho = R_\rho^{-1/2} F(L_\rho / R_\rho)^{-1} R_\rho^{-1/2},$$

where $F : \mathbb{R}^+ \to \mathbb{R}$ is an operator monotone function, which is symmetric, $F(x) = xF(x^{-1})$, and normalized, $F(1) = 1$. The operators $L_\rho$ and $R_\rho$ are the left and right multiplication operators. Clearly, $J_\rho(X) = \rho^{-1}X$ if $X$ and $\rho$ commute, so that the restriction of $\lambda$ to commutative submanifolds is the Fisher metric.

EXAMPLE 2.1. Let $J_\rho$ be the symmetric logarithmic derivative, given by $J_\rho(X) = Y$, $Y\rho + \rho Y = 2X$, then the metric $\lambda$ is monotone, with $F(x) = \frac{1+x}{2}$. This metric is sometimes called the Bures metric and it is the smallest monotone Riemannian metric.

EXAMPLE 2.2. The largest monotone metric is given by the operator monotone function $F(x) = \frac{2x}{1+x}$. In this case $J_\rho(X) = \frac{1}{2}(\rho^{-1}X + X\rho^{-1})$ is the right logarithmic derivative (RLD).

EXAMPLE 2.3. An important example of a monotone metric is the Kubo–Mori–Bogoljubov (BKM) metric, obtained from

$$\frac{\partial^2}{\partial s \partial t} \operatorname{Tr} (\rho + sX) \log(\rho + tY)|_{s,t=0} = \lambda_\rho(X, Y)$$

In this case $F(x) = \frac{x-1}{\log(x)}$.

## 3. The $\alpha$-representation and $\alpha$-connections

Let $f : \mathbb{R} \to \mathbb{R}$ be a monotone function and let $\rho \in \mathcal{M}$. Let us define the operator $L_f[\rho] : \mathcal{M}_h \to \mathcal{M}_h$ by

$$L_f[\rho](X) = \frac{d}{ds} f(\rho + sX)|_{s=0}.$$

This operator has the following properties [16]:

(i) The chain rule: $L_{f\circ g}[\rho] = L_f[g(\rho)]L_g[\rho]$. In particular, if $f$ is invertible then $L_f[\rho]$ is invertible and $L_f[\rho]^{-1} = L_{f^{-1}}[f(\rho)]$.

(ii) $L_f[\rho]$ is a self-adjoint operator in $\mathcal{M}_h$, with respect to the Hilbert–Schmidt inner product $\langle X, Y\rangle = \operatorname{Tr} X^*Y$.

(iii) If $X\rho = \rho X$, then $L_f[\rho](X) = f'(\rho)X$, $f'(x) = \frac{d}{dx} f(x)$.

Let now $f_\alpha$ be given by (1). The map

$$\ell_\alpha : \mathcal{M} \ni \rho \mapsto f_\alpha(\rho) \in \mathcal{M}_h$$

will be called the $\alpha$-embedding of $\mathcal{M}$. The $\alpha$-embedding induces the map

$$T_\rho \ni X \mapsto \mathcal{X}(f_\alpha(\rho)) = L_\alpha[\rho](X) \in \mathcal{M}_h,$$

where $L_\alpha[\rho] := L_{f_\alpha}[\rho]$, it will be called the $\alpha$-representation of the tangent vector $X$. We will often omit the indication of the point in the square brackets, if no confusion is possible.

Let $\lambda$ be a monotone metric and let $Y_1 = L_\alpha(X_1)$ and $Y_2 = L_\alpha(X_2)$ be the $\alpha$-representations of the tangent vectors $X_1$ and $X_2$, then

$$\lambda_\rho(X_1, X_2) = \operatorname{Tr} Y_1 K_\alpha(Y_2), \tag{3}$$

where $K_\alpha = L_\alpha^{-1} J_\rho L_\alpha^{-1}$.

EXAMPLE 3.1. The family of Wigner–Yanase–Dyson (WYD) metrics $\lambda_\alpha$ is defined by $J_\rho = L_{-\alpha}L_\alpha$. In [15], it was shown that such metrics are monotone for $\alpha \in [-3, 3]$ and that there are no other monotone metrics, satisfying

$$\lambda_\rho(X, Y) = \frac{\partial^2}{\partial s \partial t} \operatorname{Tr} f(\rho + sX) f^*(\rho + tY)|_{s,t=0}$$

for some functions $f$ and $f^*$. The corresponding operator monotone function is

$$F_\alpha(x) = \frac{1-\alpha^2}{4} \frac{(x-1)^2}{(x^{\frac{1+\alpha}{2}} - 1)(x^{\frac{1-\alpha}{2}} - 1)}.$$

As special cases we obtain the BKM metric for $\alpha = \pm 1$ and RLD metric for $\alpha = \pm 3$. The smallest metric in this class is the Wigner–Yanase (WY) metric, corresponding to $\alpha = 0$, here $F_0(x) = \frac{1}{4}(1+\sqrt{(x)})^2$, the Bures metric is not included. For the metric $\lambda_\alpha$, $\alpha \in [-3,3]$, we have $K_\alpha = L_{-\alpha}L_\alpha^{-1}$. It can be shown that $K_\alpha^{-1} = K_{-\alpha}$ if and only if $\lambda = \lambda_\alpha$.

The connection $\nabla^{(\alpha)}$ is defined by

$$L_\alpha((\nabla_{\mathcal{X}}^{(\alpha)}\mathcal{Y})(\rho)) = \mathcal{X}\mathcal{Y}f_\alpha(\rho)$$

for smooth vector fields $\mathcal{X}, \mathcal{Y}$. Clearly, a vector field is parallel with respect to this connection if and only if its $\alpha$-representation is a constant hermitian matrix-valued function on $\mathcal{M}$. For $\alpha = -1$ and $\alpha = 1$, we get the mixture and exponential connections, sometimes denoted by $\nabla^{(m)}$ and $\nabla^{(e)}$. The mixture connection coincides with the natural flat affine structure inherited from $\mathcal{M}_h$.

For each $\alpha$ there is a coordinate system $\xi_1, \dots, \xi_N$, such that $f_\alpha(\rho(\xi)) = \sum_i \xi_i Z_i$, where $Z_i \in \mathcal{M}_h$, $i = 1, \dots, N$, form a basis of $\mathcal{M}_h$. Clearly, such coordinate system is $\nabla^{(\alpha)}$-affine. The existence of an affine coordinate system is equivalent to flatness of the connection $\nabla^{(\alpha)}$, that is, the connections are torsion-free and the Riemannian curvature tensor vanishes. Thus we have a one-parameter family of flat $\alpha$-connections, just as in the classical case. But, contrary to the classical case, the $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ are not dual for a general monotone metric.

Let us define the connection $\nabla^{(\alpha)*}$ by

$$L_\alpha^{-1}J_\rho((\nabla_{\mathcal{X}}^{(\alpha)*}\mathcal{Y})(\rho)) = \mathcal{X}L_\alpha^{-1}J_\rho(Y) = \mathcal{X}K_\alpha L_\alpha(Y).$$

It can be easily seen from (3) that the connections $\nabla^{(\alpha)}$ and $\nabla^{(\alpha)*}$ are dual with respect to $\lambda$. It follows that $\nabla^{(\alpha)*}$ is also curvature free and it is torsion-free if and only if [16]

$$\mathcal{X}L_\alpha^{-1}J_\rho(Y) = \mathcal{Y}L_\alpha^{-1}J_\rho(X) \tag{4}$$

for all vector fields satisfying $[\mathcal{X}, \mathcal{Y}] = 0$.

THEOREM 3.1 ([17]). *Let $\alpha \in [-3,3]$. The following conditions are equivalent:*

(i) $(\nabla^{(\alpha)})^*$ *is torsion-free.*
(ii) $J_\rho = L_\alpha L_{-\alpha}$.
(iii) $(\nabla^{(\alpha)})^* = \nabla^{(-\alpha)}$.

*Proof*: (i)⇒(ii): Let $\theta \mapsto \rho(\theta)$ be a smooth parametrization of $\mathcal{M}$ and let $\partial_i = \frac{\partial}{\partial\theta_i}$, $i = 1, \dots, N$. Let us denote $X_i(\theta) = \partial_i(\rho(\theta))$. Let $\nabla^{(\alpha)*}$ be torsion-free and let $F_i(\theta) = L_\alpha^{-1}J_{\rho(\theta)}(X_i(\theta))$, $i = 1, \dots, N$. Then we get from (4) that $\partial_j F_i = \partial_i F_j$ for all $i, j$.

Let $A_1, \dots, A_N$ be a basis of $\mathcal{M}_h$ and let $F_i(\theta) = \sum_k f_{ik}(\theta)A_k$, then $\partial_i f_{jk}(\theta) = \partial_j f_{ik}(\theta)$ for all $k$, $i$ and $j$. This implies the existence of functions $\phi_1, \dots, \phi_N$, such that $f_{ik}(\theta) = \partial_i\phi_k(\theta)$. Let $\phi(\theta) = \sum_k \phi_k(\theta)A_k$, then $F_i = \partial_i\phi$. Moreover, if $\rho_t = \rho(\theta(t))$ is a curve in $\mathcal{M}$, then

$$\frac{d}{dt}\phi(\theta(t)) = \sum_i \frac{d}{dt}\theta_i(t)F_i(\theta(t)) = L_\alpha^{-1}[\rho_t]J_{\rho_t}\left(\frac{d}{dt}\rho_t\right).$$

Let now $\rho \in \mathcal{M}$ and let us consider the curve $\rho_t = \rho(\theta(t)) = t\rho + (1-t)$. Using the fact that $\frac{d}{dt}\rho_t = \rho - 1$ and $\rho_t$ commute for all $t$, we have

$$\phi(\theta(1)) - \phi(\theta(0)) = \int_0^1 \frac{d}{dt}\phi(\theta(t))dt = \int_0^1 L_\alpha^{-1}[\rho_t]J_{\rho_t}(\rho-1)dt$$
$$= \int_0^1 (1+t(\rho-1))^{\frac{\alpha-1}{2}}(\rho-1)dt = f_{-\alpha}(\rho) - f_{-\alpha}(I).$$

Therefore, $\phi(\theta) = f_{-\alpha}(\rho(\theta)) + c$. It follows that

$$L_\alpha^{-1}J_{\rho(\theta)}(X_i(\theta)) = F_i(\theta) = \partial_i f_{-\alpha}(\rho(\theta)) = L_{-\alpha}(X_i(\theta))$$

and $J_\rho = L_\alpha L_{-\alpha}$.

(ii) $\Rightarrow$ (iii) and (iii) $\Rightarrow$ (i) are quite clear. □

The statement for $\alpha = \pm 1$ was already proved in [3]. The equivalence (ii) $\Longleftrightarrow$ (iii) was proved (by a different method) in [11] for $\alpha = \pm 1$ and in [10] for $\alpha \in (-1, 1)$.

REMARK 3.1. Let $\mathcal{D} = \{\rho \in \mathcal{M} : \operatorname{Tr}\rho = 1\}$ be the submanifold of quantum states. The connections induced on $\mathcal{D}$ are orthogonal projections of the above connections. The Riemannian curvature is given by [16]

$$R_\rho^\alpha(X,Y,Z,W) = \frac{1-\alpha^2}{4}\{\operatorname{Tr} Y J_\alpha(Z)\operatorname{Tr} X J_\rho(W) - \operatorname{Tr} X J_\alpha(Z)\operatorname{Tr} Y J_\rho(W)\},$$

where $\rho \in \mathcal{D}$, $X, Y, Z, W \in T_\rho(\mathcal{D})$, and thus $R^\alpha = 0$ if and only if $\alpha = \pm 1$. Therefore, the $\alpha$-connections are not flat on $\mathcal{D}$, unless $\alpha = \pm 1$, which corresponds to the classical results.

## 4. Relative $g$-entropies and monotone metrics

Let $G$ be the set of all operator convex functions $(0,\infty) \to \mathbb{R}$, satisfying $g(1) = 0$ and $g''(1) = 1$. For $g \in G$, we define the relative $g$-entropy $H_g : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ by [20]

$$H_g(\rho,\sigma) = \operatorname{Tr}\rho^{1/2} g(L_\sigma/R_\rho)(\rho^{1/2}).$$

The set $G$ is the set of functions of the form

$$g(u) = a(u-1) + \int_{[0,\infty]} (u-1)^2\frac{1+s}{u+s}d\mu(s), \tag{5}$$

where $\mu$ is a positive finite measure on $[0,\infty]$ satisfying $\int_{[0,\infty]} d\mu(s) = 1/2$ and $a = g'(1)$ is a real number. We will denote by $b = \mu(\{\infty\})$ and $c = \mu(\{0\})$ the possible atoms in 0 and $\infty$, then

$$g(u) = a(u-1) + b(u-1)^2 + c\frac{(u-1)^2}{u} + \int_0^\infty (u-1)^2\frac{1+s}{u+s}d\mu(s).$$

For an operator convex function $g$ we define its transpose $\hat{g}(u) = ug(u^{-1})$. Clearly, $g \in G$ implies $\hat{g} \in G$, with the positive measure $\hat{\mu}$ satisfying $d\hat{\mu}(s) = d\mu(s^{-1})$ and $\hat{a} = -a$. We say that $g$ is symmetric if $g = \hat{g}$. For each symmetric function $h \in G$, we denote by $G_h \subset G$ the convex subset of functions such that $g + \hat{g} = 2h$. If $g \in G_h$, then clearly $\hat{g} \in G_h$ and $H_{\hat{g}}(\rho, \sigma) = H_g(\sigma, \rho)$.

THEOREM 4.1 ([18]). *For each* $\rho, \sigma \in \mathcal{M}$,

$$\begin{aligned} H_g(\rho,\sigma) &= a\mathrm{Tr}\,(\sigma-\rho) \\ &\quad + \mathrm{Tr}\,(\sigma-\rho)\left\{b\rho^{-1} + c\sigma^{-1} + \int_0^\infty \frac{1+s}{L_\sigma + sR_\rho}d\mu(s)\right\}(\sigma-\rho) \\ &= a\mathrm{Tr}\,(\sigma-\rho) + \mathrm{Tr}\,(\sigma-\rho)R_\rho^{-1}k(L_\sigma/R_\rho)(\sigma-\rho), \end{aligned}$$

*where*

$$k(u) = \int_0^\infty \frac{1+s}{u+s}d\mu(s) = \frac{g(u) - a(u-1)}{(u-1)^2}. \tag{6}$$

Theorem 4.1 implies that if $a = 0$, $H_g$ is a contrast functional on $\mathcal{M}$. The value of $g'(1) = a$ does not influence the Riemannian structure and connections defined by $H_g$, so that we may also use functions with $g'(1) \neq 0$, as it is sometimes more convenient, for example $g(u) = -\log u$.

Let us consider the mixture connection $\nabla^{(m)}$ on $\mathcal{M}$. A vector field on $\mathcal{M}$ is parallel with respect to $\nabla^{(m)}$ if and only if its -1-representation is a constant $\mathcal{M}_h$-valued function over $\mathcal{M}$. In the rest of the paper, we will deal only with such vector fields. The symbol $\mathcal{X}$ will denote the vector field such that the constant value of the -1-representation is $X$, similarly $\mathcal{Y}$, etc. Note that for such vector fields we have $[\mathcal{X}, \mathcal{Y}] = 0$.

Let us define the Riemannian metric $\lambda^g$ on $\mathcal{M}$ by

$$\lambda^g_\rho(X,Y) = -\frac{\partial^2}{\partial s \partial t} H_g(\rho + sX, \rho + tY)|_{s,t=0}, \quad \forall X, Y \in T_\rho. \tag{7}$$

Then [18]

$$\lambda^g_\rho(X,Y) = \mathrm{Tr}\, X R_\rho^{-1} k_{\mathrm{sym}}(L_\rho/R_\rho)(Y),$$

with

$$k_{\mathrm{sym}}(u) = k(u) + u^{-1}k(u^{-1}) = \frac{g(u) + \hat{g}(u)}{(u-1)^2}.$$

Moreover, the function $k_{\mathrm{sym}}$ is operator monotone decreasing, hence $\lambda^g$ is a monotone metric, with $F = 1/k_{\mathrm{sym}}$ the corresponding operator monotone function. Note also that if $h$ is a fixed symmetric function in $G$, then $\lambda^g$ defines the same monotone metric for each $g \in G_h$.

Conversely, if $\lambda$ is a monotone metric with the operator monotone function $F$, then

$$h(u) = \frac{1}{2}\frac{(u-1)^2}{F(u)} \tag{8}$$

is a symmetric operator convex function with $h(1) = 0$, so that $\lambda = \lambda^h$. The condition $h''(1) = 1$ is equivalent to the normalization condition $F(1) = 1$. This gives a one-to-one correspondence between the monotone metrics and the convex sets $G_h$, with symmetric $h \in G$.

## 5. The $p$-connections

Let us fix a monotone metric $\lambda$ and let $h$ be given by (8). Let us choose some $g \in G_h$, then $\lambda = \lambda^g$. We define the affine connection $\nabla^{(g)}$ by

$$\lambda_\rho(\nabla^{(g)}_{\mathcal{X}}\mathcal{Y}, \mathcal{Z}) = -\frac{\partial^3}{\partial s \partial t \partial u} H_g(\rho + sX + tY, \rho + uZ)|_{s,t,u=0},$$

just as in the classical case. It is clear that the restriction of $\nabla^{(g)}$ to submanifolds of mutually commuting elements coincides with the classical $\alpha$-connection, with $\alpha = 2g'''(1) + 3$. In contrast with the classical case, the condition $g \in G$ leads to a restriction on $\alpha$. Indeed, we have

$$g'''(1) = -6 \int_{[0,\infty]} \frac{1}{1+s} d\mu(s).$$

From this, $0 \geq g'''(1) \geq -3$ and therefore $\alpha \in [-3, 3]$ for each $g \in G$.

PROPOSITION 5.1. *The connections $\nabla^{(g)}$ and $\nabla^{(\hat{g})}$ are dual with respect to $\lambda$. Moreover, the connections are torsion-free.*

*Proof*: We have

$$\begin{aligned}
\mathcal{X}\lambda_\rho(\mathcal{Y}, \mathcal{Z}) &= -\frac{d}{du}\frac{\partial^2}{\partial t \partial s} H_g(\rho + uX + sY, \rho + uX + tZ)|_{s,t,u=0} \\
&= -\frac{\partial^3}{\partial s \partial t \partial u} H_g(\rho + uX + sY, \rho + tZ)|_{s,t,u=0} \\
&\quad - \frac{\partial^3}{\partial s \partial t \partial u} H_{\hat{g}}(\rho + uX + tZ, \rho + sY)|_{s,t,u=0} \\
&= \lambda_\rho(\nabla^{(g)}_{\mathcal{X}}\mathcal{Y}, \mathcal{Z}) + \lambda_\rho(\mathcal{Y}, \nabla^{(\hat{g})}_{\mathcal{X}}\mathcal{Z}),
\end{aligned}$$

so that duality is proved. Moreover, as $[\mathcal{X}, \mathcal{Y}] = 0$, the connection is torsion-free if $\nabla^{(g)}_{\mathcal{X}}\mathcal{Y} - \nabla^{(g)}_{\mathcal{Y}}\mathcal{X} = 0$, which is obvious. □

If the function $g$ is symmetric, then from Proposition 5.1, $\nabla^{(g)}$ is self-dual and torsion-free, hence it is the metric connection $\bar{\nabla}$. For $g \neq \hat{g}$, let us define $g_p = pg + (1-p)\hat{g}$, then $g_p \in G_h$ for $p \in [0, 1]$ and $\hat{g}_p = g_{1-p}$. For $\lambda$ and $g$ fixed, the connection given by $g_p$ will be called the $p$-connection and denoted by $\nabla^{(p)}$. Clearly, $\nabla^{(p)}$ is a convex mixture of $\nabla^{(g)}$ and $\nabla^{(\hat{g})}$,

$$\nabla^{(p)} = p\nabla^{(g)} + (1-p)\nabla^{(\hat{g})}.$$

Thus we have a one-parameter family of torsion-free $p$-connections, satisfying $(\nabla^{(p)})^* = \nabla^{(1-p)}$. We have $\nabla^{(1/2)} = \bar{\nabla}$ for all $g \in G_h$. In the rest of this section we will investigate the Riemannian curvature of the $p$-connections.

EXAMPLE 5.1. We see from (5) that the extreme boundary of $G$ consists of functions

$$g_s(u) = \frac{1+s}{2}\frac{(u-1)^2}{u+s} \quad \text{for } s \geq 0,$$
$$g_\infty(u) = \frac{1}{2}(u-1)^2.$$

We have $\hat{g}_s = g_{s^{-1}}$ for $s > 0$ and $\hat{g}_0 = g_\infty$. In this case

$$G_{h_s} = \{g_p = pg_s + (1-p)\hat{g}_s, \quad p \in [0,1]\},$$

where $h_s = \frac{1}{2}(g_s + \hat{g}_s)$. For the corresponding metric we obtain a unique family of $p$-connections. In particular, if $s = 1$, $g_1 = h_1$ is symmetric and $G_{h_1} = \{h_1\}$. The corresponding metric is the Bures metric. Hence we see that for the Bures metric, we obtain only the metric connection, which is known to be not flat, see for example [7].

EXAMPLE 5.2. Let

$$g_\alpha(u) = \begin{cases} \dfrac{4}{1-\alpha^2}\left(\dfrac{1+u}{2} - u^{\frac{1+\alpha}{2}}\right), & \alpha \neq \pm 1, \\ -\log u, & \alpha = -1, \\ u \log u, & \alpha = +1. \end{cases}$$

Then $g_\alpha \in G$ for $\alpha \in [-3, 3]$ and $\hat{g}_\alpha = g_{-\alpha}$. The relative entropies $H_{g_\alpha}$ are (up to a linear term) the $\alpha$-divergences defined by Hasegawa in [13]. It was also proved that $\lambda^{g_\alpha} = \lambda_\alpha$, the WYD metric, and $\nabla^{g_\alpha} = \nabla^{(\alpha)}$, the $\alpha$-connection from Section 3, see also [14]. Hence, $\nabla^{(g_\alpha)}$ is flat. In particular, for $\alpha = \pm 1$ we get the BKM metric and the mixture and exponential connection. The family of $p$-connections for $g(u) = -\log(u)$ is

$$\nabla^{(p)} = p\nabla^{(m)} + (1-p)\nabla^{(e)}$$

In the classical case, this is an equivalent definition of the $\alpha$-connection, $p = (1-\alpha)/2$. In our case however, these connections are different from the $\alpha$-connections which, by Theorem 3.1, have torsion-free duals with respect to the BKM metric if and only if $\alpha = \pm 1$.

To compute the Riemannian curvature tensor of $\nabla^{(p)}$, we use the theory of statistical manifolds due to Lauritzen [2]. A statistical manifold is a triple $(M, \lambda, \tilde{D})$, where $M$ is a differentiable manifold, $\lambda$ is a metric tensor and $\tilde{D}$ is a symmetric covariant 3-tensor called the skewness.

On $M$, a class of connections is introduced by

$$\nabla_X^{(p)} Y = \bar{\nabla}_X Y - \frac{1-2p}{2} D(X, Y), \tag{9}$$

where $X, Y$ are smooth vector fields, $\bar{\nabla}$ is the metric connection and the tensor $D$ is given by $\tilde{D}(X, Y, Z) = \lambda(D(X, Y), Z)$. Such connections are torsion-free, this is equivalent to symmetry of $\tilde{D}$, resp. $D$. Moreover, $(\nabla^{(p)})^* = \nabla^{(1-p)}$. Let $R^p$ be the corresponding Riemannian curvature. The manifolds satisfying $R^p = R^{1-p}$ for all $p$ are called conjugate symmetric. It was proved in [2] that the manifold is conjugate symmetric if and only if the tensor $F = \bar{\nabla}\tilde{D}$ is symmetric. From symmetry of $\tilde{D}$, it follows that $F$ is symmetric if (and only if) it is symmetric in $X$ and $Y$. We also have that if there is some $p \neq 1/2$, such that $R^p = R^{1-p}$, then the manifold is conjugate symmetric.

Let $g \in G$, then $(\mathcal{M}, \lambda^g, \tilde{D})$, where $D(\mathcal{X}, \mathcal{Y}) = \nabla_{\mathcal{X}}^{(g)} \mathcal{Y} - \nabla_{\mathcal{X}}^{(\hat{g})} \mathcal{Y}$, is a statistical manifold. The connections defined by (9) coincide with the $p$-connections if $p \in [0, 1]$. For simplicity, we denote this manifold by $(\mathcal{M}, g)$. If $g$ is symmetric, then $\tilde{D} \equiv 0$ and $\nabla^{(p)} = \bar{\nabla}$ for all $p$; in this case, the manifold is trivially conjugate symmetric.

PROPOSITION 5.2. *Let us denote* $\bar{R} = R^{1/2}$. *Then*

$$\begin{aligned} R^p(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}) = {} & \bar{R}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}) + \frac{1-2p}{2}\{F(\mathcal{Y}, \mathcal{X}, \mathcal{Z}, \mathcal{W}) - F(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W})\} \\ & + \frac{(1-2p)^2}{4}\{\lambda(D(\mathcal{X}, \mathcal{W}), D(\mathcal{Y}, \mathcal{Z})) - \lambda(D(\mathcal{X}, \mathcal{Z}), D(\mathcal{Y}, \mathcal{W}))\}. \end{aligned}$$

*Proof*: We have $[\mathcal{X}, \mathcal{Y}] = 0$ and therefore

$$R^p(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}) = \lambda(\nabla_{\mathcal{X}}^{(p)}\nabla_{\mathcal{Y}}^{(p)}\mathcal{Z} - \nabla_{\mathcal{Y}}^{(p)}\nabla_{\mathcal{X}}^{(p)}\mathcal{Z}, \mathcal{W}).$$

Let us now recall that

$$F(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}) = \mathcal{X}\tilde{D}(\mathcal{Y}, \mathcal{Z}, \mathcal{W}) - \tilde{D}(\bar{\nabla}_{\mathcal{X}}\mathcal{Y}, \mathcal{Z}, \mathcal{W}) - \tilde{D}(\mathcal{Y}, \bar{\nabla}_{\mathcal{X}}\mathcal{Z}, \mathcal{W}) - \tilde{D}(\mathcal{Y}, \mathcal{Z}, \bar{\nabla}_{\mathcal{X}}\mathcal{W}).$$

From (9) we get

$$\begin{aligned} \nabla_{\mathcal{X}}^{(p)}\nabla_{\mathcal{Y}}^{(p)}\mathcal{Z} = {} & \bar{\nabla}_{\mathcal{X}}\bar{\nabla}_{\mathcal{Y}}\mathcal{Z} - \frac{1-2p}{2}\{\bar{\nabla}_{\mathcal{X}}D(\mathcal{Y}, \mathcal{Z}) + D(\mathcal{X}, \bar{\nabla}_{\mathcal{Y}}\mathcal{Z})\} \\ & + \frac{(1-2p)^2}{4} D(\mathcal{X}, D(\mathcal{Y}, \mathcal{Z})). \end{aligned}$$

Moreover, from self-duality of $\bar{\nabla}$,

$$\lambda(\bar{\nabla}_{\mathcal{X}}D(\mathcal{Y}, \mathcal{Z}) + D(\mathcal{X}, \bar{\nabla}_{\mathcal{Y}}\mathcal{Z}), \mathcal{W}) = \mathcal{X}\tilde{D}(\mathcal{Y}, \mathcal{Z}, \mathcal{W}) - \tilde{D}(\mathcal{Y}, \mathcal{Z}, \bar{\nabla}_{\mathcal{X}}\mathcal{W}) + \tilde{D}(\mathcal{X}, \bar{\nabla}_{\mathcal{Y}}\mathcal{Z}, \mathcal{W})$$

and

$$\lambda(D(\mathcal{X}, D(\mathcal{Y}, \mathcal{Z})), \mathcal{W}) = \tilde{D}(\mathcal{X}, D(\mathcal{Y}, \mathcal{Z}), \mathcal{W}) = \lambda(D(\mathcal{X}, \mathcal{W}), D(\mathcal{Y}, \mathcal{Z})),$$

this follows from symmetry of the tensor $\tilde{D}$. Subtracting the expression with interchanged $\mathcal{X}$ and $\mathcal{Y}$ and using symmetry of $\bar{\nabla}$ completes the proof. □

COROLLARY 5.1. *Let $g \neq \hat{g}$ and let the connection $\nabla^{(g)}$ be flat. Then the manifold $(\mathcal{M}, g)$ is conjugate symmetric. Moreover, if $R^{p_0} = 0$ for some $p_0 \in (0, 1)$ then $R^p = 0$ for all $p \in [0, 1]$.*

*Proof*: If $\nabla^{(g)}$ is flat, then also its dual $\nabla^{(\hat{g})}$ is flat, therefore $0 = R^1 = R^0$ and the manifold is conjugate symmetric. From Proposition 5.2, we see that

$$0 = \bar{R}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}) + \frac{1}{4}\{\lambda(D(\mathcal{X}, \mathcal{W}), D(\mathcal{Y}, \mathcal{Z})) - \lambda(D(\mathcal{X}, \mathcal{Z}), D(\mathcal{Y}, \mathcal{W}))\},$$

and therefore

$$R^p(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}) = p(p-1)\{\lambda(D(\mathcal{X}, \mathcal{W}), D(\mathcal{Y}, \mathcal{Z})) - \lambda(D(\mathcal{X}, \mathcal{Z}), D(\mathcal{Y}, \mathcal{W}))\}.$$

If this vanishes for some $p_0 \neq 0, 1$, then the term in brackets must be zero. □

Let $\lambda$ be the BKM metric and $g(u) = -\log(u)$, then $\nabla^{(g)} = \nabla^{(m)}$ is flat. It is known [21] that in this case, the metric connection is not flat, hence $\bar{R} = R^{1/2} \neq 0$. It follows that $p\nabla^{(m)} + (1-p)\nabla^{(e)}$ is flat if and only if $p = 0$ or $p = 1$.

## 6. Operator calculus

In the following sections, we are going to prove that the connection $\nabla^{(g)}$ is flat if and only if $\nabla^{(g)} = \nabla^{(\alpha)}$ for some $\alpha \in [-3, 3]$. To do this, we will need to compute the derivatives of functions of the form $c(L_\rho, R_\rho)$. We use the same method as in [8].

Let $c$ be a function, defined and complex analytic in a neighbourhood of $(\mathbb{R}^+)^2$ in $\mathbb{C}^2$. As the operators $L_\rho$ and $R_\rho$ commute and have the same spectrum as $\rho$, by the operator calculus we have

$$c(L_\rho, R_\rho) = \frac{1}{(2\pi i)^2} \int\int c(\xi, \eta) \frac{1}{\xi - L_\rho} \frac{1}{\eta - R_\rho} d\xi d\eta,$$

where we integrate twice around the spectrum of $\rho$. We have

$$\frac{d}{dt} c(L_{\rho+tX}, R_\rho)|_{t=0} = \frac{1}{(2\pi i)^2} \int\int c(\xi, \eta) \frac{1}{\xi - L_\rho} L_X \frac{1}{\xi - L_\rho} \frac{1}{\eta - R_\rho} d\xi d\eta,$$

$$\frac{\partial^2}{\partial s \partial t} c(L_{\rho+sX+tY}, R_\rho)|_{s,t=0} = \frac{1}{(2\pi i)^2} \int\int c(\xi, \eta) \left\{ \frac{1}{\xi - L_\rho} L_Y \frac{1}{\xi - L_\rho} L_X \frac{1}{\xi - L_\rho} + \frac{1}{\xi - L_\rho} L_X \frac{1}{\xi - L_\rho} L_Y \frac{1}{\xi - L_\rho} \right\} \frac{1}{\eta - R_\rho} d\xi d\eta,$$

$$\frac{\partial^2}{\partial s \partial t} c(L_{\rho+sX}, R_{\rho+tY})|_{s,t=0} = \frac{1}{(2\pi i)^2} \int\int c(\xi, \eta) \frac{1}{\xi - L_\rho} L_X \frac{1}{\xi - L_\rho} \times \frac{1}{\eta - R_\rho} R_Y \frac{1}{\eta - R_\rho} d\xi d\eta.$$

We express the derivatives in the form of divided differences [4]. Let us denote

$$T(x, y|z) = \frac{c(x, z) - c(y, z)}{x - y}, \tag{10}$$

$$T(z|x, y) = \frac{c(z, x) - c(z, y)}{x - y}, \tag{11}$$

$$T(x, y, z|w) = \frac{T(x, y|w) - T(y, z|w)}{x - z}, \tag{12}$$

$$T(x, y|z, w) = \frac{T(x, y|z) - T(x, y|w)}{z - w} = \frac{T(x|z, w) - T(y|z, w)}{x - y}. \tag{13}$$

Then we have:

(i) $T(x, y|z)$, $T(z|x, y)$, $T(x, y|z, w)$ are symmetric in $x, y$ and $z, w$. $T(x, y, z|w)$ is symmetric in $x, y, z$.

(ii) $T(x, x|z) = \dfrac{\partial}{\partial x} c(x, z)$ and $T(z|x, x) = \dfrac{\partial}{\partial x} c(z, x)$,

(iii) $T(x, x, z|w) = \dfrac{\partial}{\partial x} T(x, z|w)$,

(iv) $T(x, x, x|w) = \dfrac{1}{2} \dfrac{\partial^2}{\partial x^2} c(x, w)$.

Let $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$ be the spectral decomposition of $\rho$. Let $e_{ij} = |\psi_i\rangle\langle\psi_j|$, then $\{e_{ij} \mid i, j = 1, \ldots, n\}$ is a basis of $\mathcal{M}_n(\mathbb{C})$. Let $u_{ij} = L_{e_{ij}}$, $v_{ij} = R_{e_{ji}}$. Then $u_{ij} e_{kl} = \delta_{jk} e_{il}$ and $v_{ij} e_{kl} = \delta_{jl} e_{ki}$. We also have

$$L_\rho = \sum_i \lambda_i u_{ii}, \qquad R_\rho = \sum_i \lambda_i v_{ii}, \qquad c(L_\rho, R_\rho) = \sum_{i,j} c(\lambda_i, \lambda_j) u_{ii} v_{jj}.$$

Let $X = \sum_{i,j} x_{ij} e_{ij}$. Inserting this into the expressions for derivatives, we get

$$\frac{d}{dt} c(L_{\rho+tX}, R_\rho)|_{t=0} = \sum_{i,j,k} T(\lambda_i, \lambda_j | \lambda_k) x_{ij} u_{ij} v_{kk}. \tag{14}$$

Similarly,

$$\frac{\partial^2}{\partial s \partial t} c(L_{\rho+sX+tY}, R_\rho)|_{s,t=0} = \sum_{i,j,k,l} T(\lambda_i, \lambda_j, \lambda_k | \lambda_l)(x_{ij} y_{jk} + y_{ij} x_{jk}) u_{ik} v_{ll}, \tag{15}$$

$$\frac{\partial^2}{\partial s \partial t} c(L_{\rho+sX}, R_{\rho+tY})|_{s,t=0} = \sum_{i,j,k,l} T(\lambda_i, \lambda_j | \lambda_k, \lambda_l) x_{ij} y_{lk} u_{ij} v_{kl}. \tag{16}$$

## 7. Conjugate symmetry

Let $g \in G$ and let $k$ be given by (6). Let us define the function $c : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ by $c(x, y) = (1/y)k(x/y)$. As we see from the integral representation, the function $k$ is operator monotone decreasing, therefore it has an analytic extension to the right halfplane in $\mathbb{C}$. It follows that $c$ is complex analytic in a neighbourhood of $(\mathbb{R}^+)^2$ and we may use the results of the previous section. Note also that for $\hat{g}$, $\hat{k}(u) = u^{-1}k(u^{-1})$ and $\hat{c}(x, y) = c(y, x)$. Moreover, $\bar{c}(x, y) = c(x, y) + \hat{c}(x, y) = (1/y)k_{\mathrm{sym}}(x/y)$ is the Chentsov–Morozova function. As follows from (6), $k(1) = \int_{[0,\infty]} d\mu = 1/2$, therefore $c(x, x) = \frac{1}{2x}$ for all $g$.

LEMMA 7.1. *Let $c$ and $\hat{c}$ be as above. Then*

$$\lambda_\rho(\nabla^{(g)}_{\mathcal{X}}\mathcal{Y}, \mathcal{Z}) = 2\mathrm{Re}\frac{d}{ds}\mathrm{Tr}\left\{X\hat{c}(L_{\rho+sY}, R_\rho)(Z) + Y\hat{c}(L_{\rho+sX}, R_\rho)(Z) - \right.$$
$$\left. - Xc(L_{\rho+sZ}, R_\rho)(Y)\right\}\Big|_{s=0}.$$

*Proof*: From Theorem 4.1 we compute

$$\begin{aligned}
\lambda_\rho(\nabla^{(g)}_{\mathcal{X}}\mathcal{Y}, \mathcal{Z}) &= -\frac{\partial^3}{\partial s\partial t\partial u}\mathrm{Tr}\,(uZ - sX - tY)c(L_{\rho+uZ}, R_{\rho+sX+tY})(uZ - sX - tY)|_{s,t,u=0}\\
&= -\frac{d}{ds}\mathrm{Tr}\left\{Xc(L_{\rho+sZ}, R_\rho)(Y) + Yc(L_{\rho+sZ}, R_\rho)(X)\right.\\
&\qquad - Xc(L_\rho, R_{\rho+sY})(Z) - Zc(L_\rho, R_{\rho+sY})(X)\\
&\qquad \left. - Yc(L_\rho, R_{\rho+sX})(Z) - Zc(L_\rho, R_{\rho+sX})(Y)\right\}\Big|_{s=0}.
\end{aligned}$$

For $\sigma, \rho \in \mathcal{M}$, $c(L_\sigma, R_\rho)$ is a positive operator on $\mathcal{M}_n(\mathbb{C})$ endowed with the inner product $\langle A, B\rangle = \mathrm{Tr}\,A^*B$. For hermitian $X$ and $Y$ we have

$$\mathrm{Tr}\,Xc(L_\sigma, R_\rho)(Y) + \mathrm{Tr}\,Yc(L_\sigma, R_\rho)(X) = 2\mathrm{Re}\mathrm{Tr}\,Xc(L_\sigma, R_\rho)(Y).$$

Clearly, for all $X \in \mathcal{M}_h$ and sufficiently small $s$, $\rho + sX \in \mathcal{M}$. Moreover,

$$\begin{aligned}
\mathrm{Re}\ \mathrm{Tr}\,Xc(L_\rho, R_{\rho+sY})(Z) &= \mathrm{Re}\ \mathrm{Tr}\,(Xc(L_\rho, R_{\rho+sY})(Z))^*\\
&= \mathrm{Re}\ \mathrm{Tr}\,X\hat{c}(L_{\rho+sY}, R_\rho)(Z). \qquad \square
\end{aligned}$$

LEMMA 7.2. *Let $D(\mathcal{X}, \mathcal{Y}) = \nabla^{(g)}_{\mathcal{X}}\mathcal{Y} - \nabla^{(\hat{g})}_{\mathcal{X}}\mathcal{Y}$ and let $\tilde{D}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) = \lambda(D(\mathcal{X}, \mathcal{Y}), \mathcal{Z})$. Let us denote $c_r(x, y) = \hat{c}(x, y) - c(x, y) = c(y, x) - c(x, y)$ and let*

$$Q(X, Y, Z) = \frac{d}{ds}\mathrm{Tr}\,Xc_r(L_{\rho+sY}, R_\rho)(Z).$$

*Then*

$$\tilde{D}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) = 2\mathrm{Re}\{Q(X, Y, Z) + Q(Y, X, Z) + Q(X, Z, Y)\} = 6Q_{\mathrm{sym}}(X, Y, Z),$$

*where $Q_{\mathrm{sym}}$ is the symmetrization of $Q$ over $X$, $Y$, $Z$.*

*Proof*: Straightforward from Lemma 7.1. □

Let us now denote by $\bar{T}(x, y|z)$ resp. $R(x, y|z)$, etc. the expressions (10)–(13) for $c = \bar{c}$, resp. $c = c_r$. From Section 6 we find

$$Q(X, Y, Z) = \sum_{i,j,k} R(\lambda_i, \lambda_j|\lambda_k) x_{ki} y_{ij} z_{jk}. \tag{17}$$

Further,

$$\begin{aligned} \mathcal{X}Q(Y, Z, W) = &\sum_{i,j,k,l} R(\lambda_i, \lambda_j, \lambda_k|\lambda_l)(x_{ij}z_{jk} + z_{ij}x_{jk}) w_{kl} y_{li} \\ &+ \sum_{i,j,k,l} R(\lambda_i, \lambda_j|\lambda_k, \lambda_l) z_{ij} w_{jl} x_{lk} y_{ki}. \end{aligned} \tag{18}$$

Clearly, $\mathcal{X}\tilde{D}(\mathcal{Y}, \mathcal{Z}, \mathcal{W})$ is the symmetrization of (18) over $\mathcal{Y}, \mathcal{Z}, \mathcal{W}$.

PROPOSITION 7.1. *Let*

$$S(x, y|z) = \frac{1}{2\bar{c}(x, y)}\{\bar{T}(x, z|y) + \bar{T}(y, z|x) - \bar{T}(x, y|z)\}.$$

*Then the -1-representation* $\bar{\nabla}_{\mathcal{X}}\mathcal{Y}(\rho) = \sum_{\alpha,\beta} d_{\alpha\beta} e_{\alpha\beta}$, *where*

$$d_{\alpha\beta} = \sum_i S(\lambda_\alpha, \lambda_\beta|\lambda_i)(x_{\alpha i} y_{i\beta} + y_{\alpha i} x_{i\beta}).$$

*Proof*: Let $h = \frac{1}{2}(g + \hat{g})$, then $\bar{\nabla} = \nabla^{(h)}$. In this case $c = \frac{1}{2}\bar{c} = \hat{c}$. From Lemma 7.1 and Eq. (14) we see that

$$\lambda_\rho(\bar{\nabla}_{\mathcal{X}}\mathcal{Y}(\rho), Z) = \operatorname{Re}\sum_{i,j,k} \bar{T}(\lambda_i, \lambda_j|\lambda_k)\{x_{ki} y_{ij} z_{jk} + y_{ki} x_{ij} z_{jk} - x_{ki} z_{ij} y_{jk}\}. \tag{19}$$

Let us denote $f^1_{\alpha\alpha} = e_{\alpha\alpha}$, for $\alpha = 1, \dots n$, $f^2_{\alpha\beta} = e_{\alpha\beta} + e_{\beta\alpha}$, $\alpha \neq \beta$, and $f^3_{\alpha\beta} = i(e_{\alpha\beta} - e_{\beta\alpha})$, $\alpha \neq \beta$. Then $\{f^1_{\alpha\alpha}, \alpha = 1, \dots, n, f^k_{\alpha\beta}, \ k = 2, 3, \alpha < \beta = 2, \dots, n\}$ forms a basis of $T_\rho$ with elements mutually orthogonal with respect to each monotone metric $\lambda$. Moreover,

$$\lambda(f^k_{\alpha\beta}, f^k_{\alpha\beta}) = \begin{cases} \bar{c}(\lambda_\alpha, \lambda_\alpha), & k = 1, \\ 2\bar{c}(\lambda_\alpha, \lambda_\beta), & k \neq 1. \end{cases}$$

Suppose that

$$\bar{\nabla}_{\mathcal{X}}\mathcal{Y}(\rho) = \sum_{k;\alpha\leq\beta} a^k_{\alpha\beta} f^k_{\alpha\beta},$$

then $\bar{\nabla}_{\mathcal{X}}\mathcal{Y}(\rho) = \sum_{\alpha,\beta} d_{\alpha\beta}e_{\alpha\beta}$, where $d_{\alpha\alpha} = a^1_{\alpha\alpha}$, $d_{\alpha\beta} = a^2_{\alpha\beta} + ia^3_{\alpha\beta}$, if $\alpha < \beta$ and $d_{\beta\alpha} = \bar{d}_{\alpha\beta}$. From (19) we compute

$$a^1_{\alpha\alpha} = 2\mathrm{Re}\sum_j S(\lambda_\alpha, \lambda_\alpha|\lambda_j)x_{\alpha j}y_{j\alpha},$$
$$a^2_{\alpha\beta} = \mathrm{Re}\sum_j S(\lambda_\alpha, \lambda_\beta|\lambda_j)\{x_{\alpha j}y_{j\beta} + y_{\alpha j}x_{j\beta}\},$$
$$a^3_{\alpha\beta} = \mathrm{Im}\sum_j S(\lambda_\alpha, \lambda_\beta|\lambda_j)\{x_{\alpha j}y_{j\beta} + y_{\alpha j}x_{j\beta}\}. \qquad \square$$

As we know from Section 5, $(\mathcal{M}, g)$ is conjugate symmetric if and only if

$$\begin{aligned}&\mathcal{X}\tilde{D}(\mathcal{Y}, \mathcal{Z}, \mathcal{W}) - \mathcal{Y}\tilde{D}(\mathcal{X}, \mathcal{Z}, \mathcal{W}) + \tilde{D}(\mathcal{X}, \bar{\nabla}_{\mathcal{Y}}\mathcal{Z}, \mathcal{W})\\ &+ \tilde{D}(\mathcal{X}, \mathcal{Z}, \bar{\nabla}_{\mathcal{Y}}\mathcal{W}) - \tilde{D}(\mathcal{Y}, \bar{\nabla}_{\mathcal{X}}\mathcal{Z}, \mathcal{W}) - \tilde{D}(\mathcal{Y}, \mathcal{Z}, \bar{\nabla}_{\mathcal{X}}\mathcal{W}) = 0.\end{aligned} \tag{20}$$

Using Lemma 7.2, (17), (18) and Proposition 7.1, we express the above equality in terms of the divided differences and then insert the basis elements $f^k_{\alpha\beta}$. This and other further lengthy computations are best performed using some software suitable for symbolic calculations, like Maple or Mathematica.

The equalities $\bar{c}(x, y) = \bar{c}(y, x)$, $c_r(x, y) = -c_r(y, x)$ and the definition and properties of divided differences imply that

$$R(x, y|x) = \frac{1}{x-y}c_r(x, y) = R(x, y|y), \tag{21}$$
$$R(x, x|x) = -\frac{\alpha}{6x^2}, \quad \text{where } \alpha = 2g'''(1) + 3, \tag{22}$$
$$R(x, y|z, w) = -R(z, w|x, y), \tag{23}$$
$$\bar{T}(x, y|z, w) = \bar{T}(z, w|x, y), \tag{24}$$
$$S(x, y|x) = \frac{1}{2}\frac{\partial}{\partial x}\log\bar{c}(x, y), \tag{25}$$
$$S(x, x|y) = \frac{1}{2}\left\{2\frac{1 - x\bar{c}(x, y)}{x-y} - x\frac{\partial}{\partial x}\bar{c}(x, y)\right\}, \tag{26}$$
$$S(x, x|x) = -\frac{1}{4x} \tag{27}$$

for all $x, y, z, w > 0$.

THEOREM 7.1. *Let $g \neq \hat{g}$ and let $\bar{g} = g + \hat{g}$, $g_r = \hat{g} - g$. If $(\mathcal{M}, g)$ is conjugate symmetric, then*

$$-\alpha\bar{g}(u) = 2ug_r'(u) - g_r(u) + 2au + 2a \tag{28}$$

*for all $u > 0$, where $a = g'(1)$ and $\alpha = 2g'''(1) + 3$.*

*Proof*: Let us write the equality (20) for the basis elements $f^k_{\alpha\beta}$ with $\alpha, \beta \in \{1, 2\}$, in this case, the resulting expression depends only on eigenvalues $\lambda_1$ and $\lambda_2$ of $\rho$. Let us put $X = Z = e_{11}$ and $Y = W = e_{12} + e_{21}$, and let $\lambda_1 = x$, $\lambda_2 = y$. We get

$$R(x,x,x|y) - R(x,x,y|x) + R(x,y|x,x) - R(x,x|x,y) \\ + 3R(x,x|x)S(x,x|y) - S(x,x|x)(2R(x,y|x) + R(x,x|y)) = 0.$$

We have

$$c_r(x,y) = \frac{y g_r(x/y)}{(x-y)^2} + \frac{2a}{x-y},$$
$$\bar{c}(x,y) = \frac{y\bar{g}(x/y)}{(x-y)^2}.$$

From this and from (i)–(iv) and (21)–(27), we get the equation

$$2g_r''\left(\frac{x}{y}\right)\frac{x}{y} + 2a + \alpha\bar{g}'\left(\frac{x}{y}\right) + g_r'\left(\frac{x}{y}\right) = 0.$$

Putting $u = x/y$ and integrating this, taking into account that $\bar{g}(1) = 0$, $g_r(1) = 0$ and $g_r'(1) = -2a$, we get (28). □

REMARK 7.1. Let $g \neq \hat{g}$, $\alpha$ and $a$ be as above. According to Theorem 7.1, if $(\mathcal{M}, g)$ is conjugate symmetric, then

$$\frac{1+\alpha}{2}\bar{g}(u) = g'(u^{-1}) + ug'(u) - au - a. \tag{29}$$

If $h$ is symmetric, then $(\mathcal{M}, h)$ is, of course, conjugate symmetric. In such a case, $\alpha = a = 0$ and Eq. (29) reads

$$h(u) = h'(u^{-1}) + uh'(u),$$

which is fulfilled for all symmetric $h \in G$.

EXAMPLE 7.1. It is easily checked that (29) is satisfied for all $pg_\alpha + (1-p)g_{-\alpha}$, $p \in [0, 1]$, $\alpha \in [-3, 3]$ (as it should be). On the other hand, it is not true for $g_s$ from the extreme boundary of $G$, unless $s = 1$, which is symmetric (the Bures case), or $s = 0$, which corresponds to $g_\alpha$, $\alpha = 3$.

## 8. Flat connections

As we know from Corollary 5.1 and Proposition 5.2, the connection $\nabla^{(g)}$ is flat if and only if

(a) $(\mathcal{M}, g)$ is conjugate symmetric,

(b) $\bar{R}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}) + \frac{1}{4}\{\lambda(D(\mathcal{X}, \mathcal{W}), D(\mathcal{Y}, \mathcal{Z})) - \lambda(D(\mathcal{X}, \mathcal{Z}), D(\mathcal{Y}, \mathcal{W}))\} = 0.$

This holds also for symmetric $g$, in that case (a) is satisfied and $D = 0$.

LEMMA 8.1.

$$\bar{R}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}) = \mathcal{X}\lambda(\bar{\nabla}_{\mathcal{Y}}\mathcal{Z}, \mathcal{W}) - \mathcal{Y}\lambda(\bar{\nabla}_{\mathcal{X}}\mathcal{Z}, \mathcal{W}) + \lambda(\bar{\nabla}_{\mathcal{X}}\mathcal{Z}, \bar{\nabla}_{\mathcal{Y}}\mathcal{W}) - \lambda(\bar{\nabla}_{\mathcal{Y}}\mathcal{Z}, \bar{\nabla}_{\mathcal{X}}\mathcal{W}).$$

*Proof*: The statement is proved similarly as Proposition 5.2, using self-duality and symmetry of $\bar{\nabla}$. □

As before, we compute

$$\mathcal{X}\lambda(\bar{\nabla}_{\mathcal{Y}}\mathcal{Z}, \mathcal{W}) = \mathrm{Re}\{\mathcal{X}\bar{Q}(Y, Z, W) + \mathcal{X}\bar{Q}(Z, Y, W) - \mathcal{X}\bar{Q}(Y, W, Z)\},$$

where

$$\mathcal{X}\bar{Q}(Y, Z, W) = \sum_{i,j,k,l} \bar{T}(\lambda_i, \lambda_j, \lambda_k|\lambda_l)(x_{ij}z_{jk} + z_{ij}x_{jk})w_{kl}y_{li} + \sum_{i,j,k,l} \bar{T}(\lambda_i, \lambda_j|\lambda_k, \lambda_l)z_{ij}w_{jl}x_{lk}y_{ki}. \tag{30}$$

Moreover,

$$\lambda(X, Y) = \mathrm{Tr}\, X\bar{c}(L_\rho, R_\rho)(Y) = \sum_{i,j} \bar{c}(\lambda_i, \lambda_j)x_{ji}y_{ij}. \tag{31}$$

The second term in (b) can be written in a form using $\tilde{D}$: let $\{b_j \mid j = 1, \dots N\}$ be the orthonormal basis obtained by normalization of $\{f^k_{\alpha\beta} \mid k = 1, 2, 3, \alpha \le \beta = 1, \dots, n\}$, then

$$\begin{aligned}&\lambda(D(\mathcal{X}, \mathcal{W}), D(\mathcal{Y}, \mathcal{Z})) - \lambda(D(\mathcal{X}, \mathcal{Z}), D(\mathcal{Y}, \mathcal{W}))\\ &= \sum_j \{\tilde{D}(X, W, b_j)\tilde{D}(Y, Z, b_j) - \tilde{D}(X, Z, b_j)\tilde{D}(Y, W, b_j)\}.\end{aligned} \tag{32}$$

Using Lemma 8.1, (30), (31), (32) and Proposition 7.1, we get from (b) an equation involving divided differences, and we may proceed in the same way as in Section 6.

PROPOSITION 8.1. *Let $g \in G$. If the connection $\nabla^{(g)}$ is flat, then*

$$(\alpha^2 - 1)\bar{g}(u) + \bar{g}'(u)(u - 1) - 2\bar{g}''(u)u(1 + u) + \alpha(g_r'(u) + 2a)(u - 1) + 8 = 0 \tag{33}$$

*for all $u > 0$.*

*Proof*: Let $X = Z = e_{11}$ and $Y = W = e_{12} + e_{21}$. From (b) we get the equation

$$\begin{aligned}&2\bar{T}(x, x, x|y) - 2\bar{T}(x, x|x, y) - 2\bar{c}(x, y)S(x, y|x)^2 + 4\bar{c}(x, x)S(x, x|x)\bar{S}(x, x|y)\\ &- 3\frac{R(x, x|x)}{\bar{c}(x, x)}(2R(x, y|x) + R(x, x|y)) + \frac{1}{\bar{c}(x, y)}(2R(x, y|x) + R(x, x|y))^2 = 0.\end{aligned}$$

For $X = Z = e_{12} + e_{21}$, $Y = W = i(e_{12} - e_{21})$, Eq. (b) reads

$$4\bar{T}(y, y, x|x) + 4\bar{T}(x, x, y|y) - 8\bar{T}(x, y|x, y) + 4\bar{c}(x, x)S(x, x|y)^2$$
$$+ 4\bar{c}(y, y)S(y, y|x)^2 - (2R(x, y|x) + R(x, x|y))^2 - (2R(x, y|y) + R(y, y|x))^2 = 0.$$

As in the proof of Theorem 7.1, we get after some rearrangements

$$u\left[(g_r'(u) + 2a)^2 - (\bar{g}'(u))^2\right] + \bar{g}(u)\{2u\bar{g}''(u) + \bar{g}'(u) + \alpha(g_r'(u) + 2a)\} = 0$$

from the first equation, and

$$u\left[(g_r'(u) + 2a)^2 - (\bar{g}'(u))^2\right] + \{g_r'(u)u - g_r(u) + 2a\}^2 - \{\bar{g}'(u)u - \bar{g}(u)\}^2 + 8\bar{g}(u) = 0,$$

from the second equation.

If $g$ is symmetric, then in the above two equations $\alpha = a = 0$ and $g_r = 0$. From this we get

$$\bar{g}(u)\{-\bar{g}(u) + \bar{g}'(u)(u - 1) - 2\bar{g}''(u)u(1 + u) + 8\} = 0,$$

which is (33).

Let now $g \neq \hat{g}$. From (a), $(\mathcal{M}, g)$ is conjugate symmetric, and therefore (28) holds. From this

$$g_r'(u)u - g_r(u) + 2a = -\alpha\bar{g}(u) - u\{g_r'(u) + 2a\}.$$

Inserting this into the second equation and after some further computation we get (33). □

We are now in a position to prove our main theorem.

THEOREM 8.1. *Let $g \in G$ and $\alpha = 2g'''(1) + 3$. Then $\alpha \in [-3, 3]$ and the connection $\nabla^{(g)}$ is flat if and only if $\nabla^{(g)} = \nabla^{(\alpha)}$.*

*Proof*: Let $g$ be symmetric and suppose that $\nabla^{(g)}$ is flat. Then $\bar{g} = 2g$ and we get from (33) that $g$ is a solution of

$$-g(u) + g'(u)(u - 1) - 2g''(u)u(1 + u) + 4 = 0$$

with the initial conditions $g(1) = 0$, $g'(1) = 0$. The unique solution of this equation is

$$g(u) = 2(1 - \sqrt{u})^2 = g_0.$$

If $g \neq \hat{g}$, then from (28) and (33) we get that $g_r$ is the solution of

$$(\alpha^2 - 1)g_r(u) - (\alpha^2 - 1)(1 + u)g_r'(u) + 4u(u + 2)g_r''(u)$$
$$+ 4u^2(u + 1)g_r'''(u) - 4a(\alpha^2 - 1) + 8\alpha = 0$$

with $g_r(1) = 0$, $g_r'(1) = -2a$ and $g_r''(1) = 0$. If $\alpha \neq \pm 1$, the unique solution is

$$\begin{aligned} g_r(u) &= \frac{4}{1-\alpha^2}(u^{\frac{1+\alpha}{2}} - u^{\frac{1-\alpha}{2}}) - \left(\frac{4\alpha}{1-\alpha^2} + 2a\right)(u-1) \\ &= g_{-\alpha}(u) - g_\alpha(u) - 2\big(a - g_\alpha'(1)\big)(u-1), \end{aligned}$$

and from (28) we get $\bar{g} = g_\alpha + g_{-\alpha}$.

If $\alpha = -1$, then the solution of the above equation is

$$g_r(u) = \log(u)(u+1) - 2(a - g_{-1}'(1))(u-1),$$

and from (28) we get

$$\bar{g}(u) = \log(u)(u-1).$$

It follows that $g = g_\alpha$, up to an additional linear term $(g'(1) - g_\alpha'(1))(u-1)$. □

COROLLARY 8.1. *Let $\lambda$ be a monotone Riemannian metric and let $\bar{\nabla}$ be the metric connection. Then $\bar{\nabla}$ is flat if and only if $\lambda$ is the WY metric ($\alpha = 0$).*

*Proof*: Let $G_h$ be the convex subset of $G$, corresponding to $\lambda$. Then $\bar{\nabla} = \nabla^{(h)}$ and $h = \hat{h}$ implies that $h'''(1) = -\frac{3}{2}$. The proof now follows from Theorem 8.1. □

## Acknowledgement

The research was supported by the grant VEGA 1/0264/03.

REFERENCES

[1] S. Amari: *Differential-geometrical Methods in Statistics*, Lecture Notes in Statistics **28**, Springer, Berlin 1985.
[2] S. Amari, O. E. Barndorff-Nielsen, R. E. Kass, S. L. Lauritzen and C. R. Rao: *Differential Geometry in Statistical Inference*, IMS Lecture Notes-Monograph Series 10, Mayward 1987.
[3] S. Amari and H. Nagaoka: *Method of Information Geometry*, AMS monograph, Oxford University Press, 2000.
[4] R. Bhatia: *Matrix Analysis*, Springer, New York 1997.
[5] N. N. Chentsov: *Statistical Decision Rules and Optimal Inferences*, Translation of Math. Monog. **53**, AMS, Providence, 1982.
[6] N. N. Chentsov and Morozova: Markov invariant geometry on state manifolds, *Itogi Nauki i Tekhniki* **36** (1990), 69.
[7] J. Dittmann: *J. Geom. Phys.* **31** (1999), 16.
[8] J. Dittmann: *Lin. Alg. Appl.* **315** (2000), 83.
[9] S. Eguchi: *Ann. Statist.* **11** (1983), 793.
[10] M. R. Grasselli: Monotonicity, Duality and Uniqueness of the WYD Metrics, submitted to IDAQP, math-ph/0212022.
[11] M. R. Grasselli and R. F. Streater: IDAQP, **4** (2001), 173, math-ph/0006030
[12] H. Hasegawa: *Rep. Math. Phys.* **39** (1997), 49.
[13] H. Hasegawa: *Rep. Math. Phys.* **33** (1993), 87.
[14] H. Hasegawa: On the dual geometry of Wigner–Yanase–Dyson information contents, 2002, to appear in IDAQP.
[15] H. Hasegawa and D. Petz: Non-commutative extension of information geometry II, in: *Quantum Communication, Computing and Measurement*, Hirota et al. (eds.), Plenum Press, New York 1997.

[16] A. Jenčová: *Rep. Math. Phys.* **47** (2001), 121.
[17] A. Jenčová: Dualistic properties of the manifold of quantum states, in: *Disordered and Complex Systems*, AIP Conf. Proc. **553**, AIP, Melville, NY, 2001.
[18] A. Lesniewski and M. B. Ruskai: *J. Math. Phys.* **40** (1999), 5702, math-ph/9808016.
[19] H. Nagaoka: Differential geometrical aspects of quantum state estimation and relative entropy,in: *Quantum Communication, Computing and Measurement*, Hirota et al. (eds.), Plenum Press, New York 1994.
[20] D. Petz: *Rep. Math. Phys.* **23** (1986), 57.
[21] D. Petz: *J. Math. Phys.* **244** (1994), 780.
[22] D. Petz: *Linear Algebra Appl.* **244** (1996), 81.

Infinite Dimensional Analysis, Quantum Probability
and Related Topics
Vol. 8, No. 2 (2005) 215–233




# QUANTUM INFORMATION GEOMETRY AND NONCOMMUTATIVE $L_p$-SPACES

ANNA JENČOVÁ

*Mathematical Institute, Slovak Academy of Sciences,*
*Štefánikova 49, 81473 Bratislava, Slovakia*
*jenca@mat.savba.sk*

Received 10 March 2004
Revised 21 September 2004
Communicated by D. Petz

Let $M$ be a von Neumann algebra. We define the noncommutative extension of information geometry by embeddings of $M$ into noncommutative $L_p$-spaces. Using the geometry of uniformly convex Banach spaces and duality of the $L_p$ and $L_q$ spaces for $1/p+1/q=1$, we show that we can introduce the $\alpha$-divergence, for $\alpha \in (-1,1)$, in a similar manner as Amari in the classical case. If restricted to the positive cone, the $\alpha$-divergence belongs to the class of quasi-entropies, defined by Petz.

*Keywords*: Noncommutative $L_p$-spaces; $\alpha$-connections; $\alpha$-embeddings; duality; divergences.

AMS Subject Classification: 58B99, 53Z05

## 1. Introduction

The classical information geometry deals with the differential geometric aspects of families of probability densities with respect to a given measure $\mu$. The theory, developed in Refs. 1 and 5, has already been extended to the nonparametric case, where the manifold is modeled on some infinite dimensional Banach space, see Refs. 22 and 8.

Noncommutative version of the theory has also been proposed, mostly restricted to (invertible) density operators on finite dimensional Hilbert spaces,[11,14,17,21] but there are results in infinite dimensions.[9,12,23,24]

An interesting part of the classical (finite-dimensional) information geometry, developed by Amari and Nagaoka,[1,2] deals with the structure of Riemannian manifolds with a pair of dual flat affine connections. For such manifolds, there is a pair $(\theta, \eta)$ of dual affine coordinate systems, related by Legendre transformations

$$\theta_i = \frac{\partial}{\partial \eta_i}\varphi(\eta)\,, \qquad \eta_i = \frac{\partial}{\partial \theta_i}\psi(\theta)\,,$$

where $\psi$, $\varphi$ are potential functions, satisfying

$$\psi(\theta) + \varphi(\eta) - \sum_i \theta_i \eta_i = 0 \,.$$

A quasi-distance, called the divergence, is then defined by

$$D(\theta_1, \theta_2) = \psi(\theta_1) + \varphi(\eta_2) - \sum_i \theta_{1i} \eta_{2i} \,.$$

For manifolds of probability density functions, flat with respect to the $\pm\alpha$-connections, the corresponding $\alpha$-divergence belongs to the class of Cziszár's $f$-divergences

$$S_f(p,q) = \int f\left(\frac{q}{p}\right) dp \,,$$

where $f$ is a convex function.

The $f$-divergences were generalized to von Neumann algebras by Petz in Ref. 20 by means of the relative modular operator of normal positive functionals on $M$:

$$S_g(\phi, \psi) = (g(\Delta_{\phi,\psi})\xi_\psi, \xi_\psi) \,,$$

where $f$ is operator convex and $\xi_\psi$ is the vector representative of $\psi$. On the other hand, Amari's construction of the $\alpha$-divergence, starting from a pair of dual flat connections, was extended to the manifold of faithful positive linear functionals on a matrix algebra $\mathcal{M}_n(\mathbb{C})$,[14,11] and it was shown that this divergence belongs to the class defined by Petz. The main purpose of this paper is the extension of this construction to all von Neumann algebras.

One of the important results of the infinite dimensional, classical and quantum, version of information geometry is the definition of Amari–Chentsov $\alpha$-connections for $\alpha \in (-1,1)$ given in Refs. 8 and 9. This definition uses the $\alpha$-embedding of the manifold of density functions (or density operators with respect to a n.s.f. trace on a semifinite von Neumann algebra) into the unit sphere of the (noncommutative) $L_p$-space, with $p = 2/(1-\alpha)$. It was shown that the Amari–Nagaoka duality of the $+\alpha$ and $-\alpha$-connections is exactly the $L_p$-space duality and that the fact that these spaces for $1 < p < \infty$ are uniformly convex is basic for this definition.

In this paper, the $\alpha$-embedding is defined in a similar manner, but it is extended to the whole predual $M_*$. This embedding is used to define the manifold structure on $M_*$. The flat $\alpha$-connections are induced from the trivial connection on $L_p$, in our setting the connection is defined on the tangent bundle. Its dual, living on the cotangent bundle, is the $-\alpha$-connection. Moreover, the $\pm\alpha$-embeddings define a pair of dual coordinates on $M_*$. Using the uniform convexity of the $L_p$ spaces, it is shown that the dual coordinates are related by potential functions, just as in Amari's theory. From this, we can define a divergence functional on $L_p(M, \phi)$.

Via the $\alpha$-embedding, the divergence in $L_p(M, \phi)$ induces a functional on $M_* \times M_*$, which is called the $\alpha$-divergence. We will show that if restricted to the positive cone, the $\alpha$-divergence is exactly the Petz quasi-entropy $S_{g_\alpha}$, with

$$g_\alpha(t) := \frac{2}{1-\alpha} + \frac{2}{1+\alpha}t - \frac{4}{1-\alpha^2}t^{\frac{1+\alpha}{2}}.$$

We will further investigate the properties of the divergence in $L_p(M,\phi)$, especially the projection theorems. These imply some existence and uniqueness results for the $\alpha$-projections, which generalize the projection theorems in Ref. 1.

## 2. Geometry of Banach Spaces

In this section, we list some facts about convexity and smoothness of Banach spaces, see Refs. 15 and 7 for details.

Let $X$ be a Banach space and let $X^*$ be the dual of $X$. For $u \in X^*$ we denote $\langle x, u\rangle = u(x)$. Let $K$ be a closed convex subset in $X$ with nonempty interior and let $S$ be the boundary of $K$. In particular, let $S_r$ be the sphere with radius $r$ in $X$.

### 2.1. *Supporting hyperplanes and functionals*

Let us first recall that a (closed) hyperplane in $X$ is a linear manifold $x + H$ with $x \in X$ and $H$ a closed linear subspace of codimension 1. Each hyperplane is uniquely given by $x$ and an element $u \in X^*$, $\|u\| = 1$, having $H$ as the null-space. Conversely, each $u \in X^*$ defines the hyperplane

$$x + \mathrm{Ker}\ u = \{y \in X, \langle y, u\rangle = \gamma = \langle x, u\rangle\}.$$

If $x + H$ is a real hyperplane in a complex Banach space, then $H$ is given by a real linear continuous functional $u$ and there is a unique $v \in X^*$, such that $u = \Re v$. Each real hyperplane determines two closed half-spaces $\{\langle x, u\rangle \le \gamma\}$ and $\{\langle x, u\rangle \ge \gamma\}$.

A *supporting hyperplane* of $K$ is a real hyperplane $x + H$, containing at least one point of $K$ and such that $K$ lies in one of the two closed half-spaces determined by $x + H$. If $u$ is the corresponding real linear functional, then $u$ (or $-u$) attains its maximum on $K$. There is at least one supporting hyperplane through every boundary point of $K$.

### 2.2. *Smoothness and strict convexity*

A point $x \in S_1$ is called a *point of smoothness* if there is exactly one supporting hyperplane passing through $x$, the *tangent hyperplane* at $x$. Equivalently, there is a unique point $v_x \in X^*$, $\|v_x\| = 1$, such that $v_x$ attains its norm at $x$. The tangent hyperplane is then determined by $u = \Re v_x$ and all points in the unit ball satisfy $\Re\langle y, v_x\rangle \le 1$. If each element in $S_1$ is a point of smoothness, then we say that $X$ is *smooth*.

The norm in $X$ is said to be *weakly* (*Gateaux*) *differentiable* at $x \in S_1$ if for each $y \in S_1$, the limit

$$\lim_{t\to 0} \frac{\|x+ty\| - \|x\|}{t} =: q'(x,y) \tag{1}$$

exists. The space $X$ is smooth if and only if its norm is weakly differentiable at each $x \in S_1$. The weak derivative is given by

$$q'(x, y) = \Re\langle y, v_x\rangle .$$

In such a case, the norm is weakly differentiable at each $x \in X$ except the origin, and $q'(x, y) = \Re\langle y, v_{x/\|x\|}\rangle$, where $v_{x/\|x\|}$ is the unique point in the unit sphere $S_1^*$ in $X^*$, such that $\langle x, v_{x/\|x\|}\rangle = \|x\|$.

We say that $K$ is *strictly convex*, if every boundary point of $K$ is an extreme point, that is, if $S$ contains no line segment. In this case, every supporting hyperplane of $K$ contains exactly one point of $S$. The space $X$ is strictly convex if the closed unit ball $K_1$ is strictly convex. There is a duality between strict convexity and smoothness of a Banach space:

(i) If $X^*$ is strictly convex, then $X$ is smooth.
(ii) If $X^*$ is smooth, then $X$ is strictly convex.

### 2.3. *Uniform smoothness and uniform convexity*

We say that the norm in $X$ is *strongly (Frèchet) differentiable* at $x \in S_1$ if the limit (1) exists uniformly for $y \in S_1$, if this is true for all $x \in S_1$, the norm is termed *strongly differentiable.* The norm is *uniformly strongly differentiable* if (1) exists uniformly in $x$, $y \in S_1$. Clearly, if the norm is strongly differentiable, then $X$ is smooth and the map

$$S_1 \ni x \mapsto v_x \in S_1^*$$

is well defined. Moreover, we have

**Theorem 2.1.**[7,6] *The norm is (uniformly) strongly differentiable if and only if the map $x \mapsto v_x$ is single-valued and norm to norm (uniformly) continuous from $S_1$ to $S_1^*$.*

The space $X$ is *uniformly smooth* if for each $\varepsilon > 0$ there is an $\eta(\varepsilon) > 0$, such that $\|x\| \geq 1$, $\|y\| \geq 1$ and $\|x-y\| \leq \eta(\varepsilon)$ always implies $\|x+y\| \geq \|x\| + \|y\| - \varepsilon\|x-y\|$.

The dual notion to uniform smoothness is uniform convexity: $X$ is *uniformly convex* if for each $0 < \varepsilon \leq 2$ there is a $\delta(\varepsilon) > 0$ such that if $x$, $y \in K_1$ and $\|x - y\| \geq \varepsilon$, then $\|\frac{1}{2}(x + y)\| \leq 1 - \delta(\varepsilon)$. The function $\delta(\varepsilon)$ is called the *module of convexity.* Every uniformly convex Banach space is strictly convex and reflexive. Moreover, the following statements are equivalent:

(i) The norm on $X$ ($X^*$) is uniformly strongly differentiable.
(ii) $X$ is uniformly smooth (uniformly convex).
(iii) $X^*$ is uniformly convex (uniformly smooth).

Let us now define the map $F : X \setminus \{0\} \to X^* \setminus \{0\}$ by

$$F(x) = \|x\|v_{x/\|x\|} \, .$$

Then $F$ is a *support mapping*,[7]

(i) $\|x\| = 1$ implies $\|F(x)\| = 1 = \langle x, F(x)\rangle$,
(ii) $\lambda \geq 0$ implies $F(\lambda x) = \lambda F(x)$.

It follows from Sec. 2.2 that if $X$ is smooth, the support mapping is unique. If we put $F(0) = 0$, then we have

**Theorem 2.2.**[6] *Let the Banach space $X$ be uniformly convex and let the norm be strongly differentiable. Then $F$ is a homeomorphism of $X$ onto $X^*$ (in the norm topologies).*

## 3. Noncommutative $L_p$ Spaces

In this section, we recall the definition and some properties of noncommutative $L_p$ spaces on a von Neumann algebra, following the approach in Refs. 4 and 16.

Let $M$ be a von Neumann algebra and let $\phi$ be a faithful normal semifinite weight. We denote $N_\phi$ the set of $y \in M$ satisfying $\phi(y^*y) < \infty$ and $M_0$ the set of all elements in $N_\phi \cap N_\phi^*$, entire analytic with respect to the modular automorphism $\sigma_t^\phi$ associated with $\phi$. We also denote the GNS map by $N_\phi \ni y \mapsto \eta_\phi(y) \in H_\phi$.

Let $1 \leq p \leq \infty$. The noncommutative $L_p$-space with respect to $\phi$ is the space $L_p(M, \phi)$ of all closed operators acting on the Hilbert space $H_\phi$, satisfying

$$TJ_\phi \sigma^\phi_{-i/p}(y) J_\phi \supset J_\phi y J_\phi T\,,$$

for all $y \in M_0$, such that the $L_p$-norm

$$\|T\|_p = \left\{ \sup_{x \in M_0, \|x\| \leq 1} \| |T|^{p/2} \eta_\phi(x)\| \right\}^{2/p}$$

is finite. Then $L_p(M, \phi)$ with this norm is a Banach space. Let $1 < p < \infty$, then $L_p(M, \phi)$ is uniformly convex and uniformly strongly differentiable. The dual space $L_p^*(M, \phi)$ is $L_q(M, \phi)$, with $1/p + 1/q = 1$, where the duality is given by

$$\langle T, T'\rangle_\phi = \lim_{y \to 1} (T\eta_\phi(y), T'\eta_\phi(y))\,, \tag{2}$$

where $T \in L_p(M, \phi)$, $T' \in L_q(M, \phi)$. The limit is taken in the $*$-strong topology with restriction $y \in M_0$, $\|y\| \leq 1$. Each $T \in L_p(M, \phi)$, $1 \leq p < \infty$, has a unique polar decomposition of the form

$$T = u\Delta_{\psi,\phi}^{1/p}\,, \tag{3}$$

where $\psi \in M_*^+$, $u \in M$ is a partial isometry, such that the support projection $s(\psi) = u^*u$ and $\Delta_{\psi,\phi}$ is the relative modular operator, see Appendix C in Ref. 4 for definition and basic properties. We have

$$\|u\Delta_{\psi,\phi}^{1/p}\|_p = \psi(1)^{1/p}\,. \tag{4}$$

On the other hand, each operator of the form (3) is in $L_p(M,\phi)$. The positive cone $L_p^+(M,\phi)$ is the set of positive operators in $L_p(M,\phi)$ and we have

$$L_p^+(M,\phi)=\{\Delta_{\psi,\phi}^{1/p},\psi\in M_*^+\}.$$

The identity

$$\varphi(au)=\langle u\Delta_{\varphi,\phi},a^*\rangle_\phi \tag{5}$$

for $a\in M$ gives an isometric isomorphism of $M_*$ and $L_1(M,\phi)$. Similarly, $L_2(M,\phi)$ is isomorphic to $H_\phi$ by

$$u\Delta_{\varphi,\phi}^{1/2}\mapsto u\xi_\varphi,$$

where $\xi_\varphi$ is the vector representative of $\varphi$ in the natural positive cone in $H_\phi$.

If $\tilde\phi$ is a different n.s.f. weight, then there is an isometric isomorphism $\tau_p(\tilde\phi,\phi)$ : $L_p(M,\phi)\to L_p(M,\tilde\phi)$ and

$$\langle T,T'\rangle_\phi=\langle\tau_p(\tilde\phi,\phi)T\,,\tau_q(\tilde\phi,\phi)T'\rangle_{\tilde\phi} \tag{6}$$

holds for all $T\in L_p(M,\phi)$ and $T'\in L_q(M,\phi)$.

A bilinear form on $L_p(M,\phi)\times L_q(M,\phi)$ is defined by

$$[T,T']_\phi=\langle T,T'^*\rangle_\phi\,,\qquad T\in L_p(M,\phi),T'\in L_q(M,\phi)\,.$$

If $T_k\in L_{p_k}(M,\phi)$, $\sum_k 1/p_k=1/r$, then the product $T=T_1\cdots T_n$ is well defined as an element of $L_r(M,\phi)$ and

$$\|T\|_r\le\|T_1\|_{p_1}\cdots\|T_n\|_{p_n}\,.$$

If $r=1$, then

$$\begin{aligned}[][T_1\cdots T_n]_\phi:=[T,1]_\phi&=[T_1\cdots T_k,T_{k+1}\cdots T_n]_\phi\\&=[T_{k+1}\cdots T_nT_1\cdots T_k]_\phi\end{aligned} \tag{7}$$

for each $1\le k\le n-1$ and

$$|[T_1\cdots T_n]_\phi|\le\|T_1\|_{p_1}\cdots\|T_n\|_{p_n}\,. \tag{8}$$

Let $L_p^h(M,\phi)=\{T\in L_p(M,\phi),T=T^*\}$. Since the adjoint operation is a conjugate linear isometry on $L_p(M,\phi)$, $L_p^h(M,\phi)$ is a real Banach subspace of $L_p(M,\phi)$. Moreover, for $1<p<\infty$, $L_p^h(M,\phi)$ is uniformly convex and the dual $L_p^{h*}(M,\phi)=L_q^h(M,\phi)$ for $1/p+1/q=1$.

## 4. The Manifold and Dual Affine Connections

Let $M$ be a von Neumann algebra and let $\phi$ be a faithful normal semifinite weight.

For $-1<\alpha<1$, we define the noncommutative $\alpha$-embedding by

$$\ell_\alpha^\phi:M_*\to L_p(M,\phi)\,,\qquad p=\frac{2}{1-\alpha}$$

$$\omega\mapsto pu\Delta_{\varphi,\phi}^{1/p}\,,$$

where $\omega(a) = \varphi(au)$, $a \in M$ is the polar decomposition of $\omega$. It is clear from uniqueness of the polar decompositions that $\ell_\alpha^\phi$ is bijective. Moreover, it maps the hermitian (that is, $\omega(a^*) = \overline{\omega(a)}$) elements in $M_*$ onto $L_p^h(M,\phi)$ and $M_*^+$ onto the positive cone $L_p^+(M,\phi)$.

If $\psi$ is a different f.n.s. weight, then the space $L_p(M,\psi)$ is identified with $L_p(M,\phi)$ by the isometric isomorphism $\tau_p(\psi,\phi)$. The corresponding $\alpha$-embeddings are related by

$$\ell_\alpha^\psi = \tau_p(\psi,\phi)\ell_\alpha^\phi \,.$$

We denote by $\mathcal{M}_\alpha$ the set $M_*$ with the manifold structure induced from $\ell_\alpha^\phi$. Due to the above isomorphism, the manifold structure does not depend of the choice of $\phi$. For $\omega \in M_*$, $\ell_\alpha^\phi(\omega) \in L_p(M,\phi)$ will be called the $\alpha$-coordinate of $\omega$. The $-\alpha$-coordinate is an element of the dual space $L_q(M,\phi)$, $1/p + 1/q = 1$. Moreover, for $\omega_1$, $\omega_2 \in M_*$ and a n.s.f. weight $\psi$, we have by (6)

$$\begin{aligned} \langle \ell_\alpha^\psi(\omega_1), \ell_{-\alpha}^\psi(\omega_2)\rangle_\psi &= \langle \tau_p(\psi,\phi)\ell_\alpha^\phi(\omega_1), \tau_q(\psi,\phi)\ell_{-\alpha}^\phi(\omega_2)\rangle_\psi \\ &= \langle \ell_\alpha^\phi(\omega_1), \ell_{-\alpha}^\phi(\omega_2)\rangle_\phi \,. \end{aligned} \tag{9}$$

In the sequel, we will just write $\ell_\alpha$ instead of $\ell_\alpha^\phi$. We will say that $\ell_\alpha(\omega)$ and $\ell_{-\alpha}(\omega)$ are dual coordinates of $\omega \in M_*$.

**Remark 4.1.** It is clear that it is also possible to define the manifold structure on hermitian elements in $M_*$, using the real Banach space $L_p^h(M,\phi)$. All the subsequent statements hold also for this case.

The trivial connection on $L_p(M,\phi)$ induces a globally flat affine connection on the tangent bundle $T\mathcal{M}_\alpha$, called the $\alpha$-connection. Let us recall that there is a one-to-one correspondence between affine connections and parallel transports on $T\mathcal{M}_\alpha$. If the connection is globally flat, the parallel transport is given by a family of isomorphisms $U_{x,y} : T_x(\mathcal{M}_\alpha) \to T_y(\mathcal{M}_\alpha)$, $x$, $y \in \mathcal{M}_\alpha$, satisfying

(i) $U_{x,x} = \mathrm{Id}$,
(ii) $U_{y,z}U_{x,y} = U_{x,z}$.

In our case, the tangent space $T_x(\mathcal{M}_\alpha)$ can be identified with $L_p(M,\phi)$ and the map $U_{x,y}$ is the identity map for all $x$, $y \in \mathcal{M}_\alpha$. We define the dual connection as in Ref. 8, i.e. a linear connection on the cotangent bundle $T^*\mathcal{M}_\alpha$, such that the corresponding parallel transport $U^*$ satisfies

$$\langle v, U_{x,y}^*(w)\rangle_\phi = \langle U_{y,x}(v), w\rangle_\phi = \langle v, w\rangle_\phi$$

for $w \in (T_x(\mathcal{M}_\alpha))^* \equiv L_q(M,\phi)$ and $v \in T_y(\mathcal{M}_\alpha)$. Obviously, $U^*$ is the trivial parallel transport in $L_q(M,\phi)$, hence the dual of the $\alpha$-connection is the $-\alpha$-connection.

## 5. Duality

The $L_p$ spaces for $1 < p < \infty$ are uniformly convex and uniformly smooth, therefore we can use the results of Sec. 2.

Let $\omega \in M_*$. We will show how $\omega$ is related to its dual coordinates.

**Proposition 5.1.** *Let $\omega \in M_*$ and let $\omega(a) = \psi(au)$ be the polar decomposition. Then*

$$pq\omega(a) = \langle \ell_\alpha(\omega), a^* u^* \ell_{-\alpha}(\omega)\rangle_\phi\,, \qquad a \in M\,.$$

**Proof.** We have $s(\Delta_{\psi,\phi}) = s(\psi) = u^*u$. From this and from (5), (7) it follows that

$$\begin{aligned}\psi(au) &= \langle u\Delta_{\psi,\phi}, a^*\rangle_\phi = [u\Delta_{\psi,\phi}u^*ua]_\phi = [u\Delta^{1/p}_{\psi,\phi}\Delta^{1/q}_{\psi,\phi}u^*ua]_\phi \\ &= [u\Delta^{1/p}_{\psi,\phi}, \Delta^{1/q}_{\psi,\phi}u^*ua]_\phi = \frac{1}{pq}\langle \ell_\alpha(\omega), a^*u^*\ell_{-\alpha}(\omega)\rangle_\phi\,.\end{aligned}$$

□

Let $x = \ell_\alpha(\omega)$ and $\tilde{x} = \ell_{-\alpha}(\omega)$ be the dual coordinates of $\omega \in M_*$. The map

$$x \mapsto \tilde{x} = \ell_{-\alpha}\ell_\alpha^{-1}(x)$$

is called the duality map. It is easy to see from (4) and Proposition 5.1 that for $x \in L_p(M, \phi)$ we have

$$\left\|\frac{\tilde{x}}{q}\right\|_q^q = \left\|\frac{x}{p}\right\|_p^p = \frac{1}{pq}\langle x, \tilde{x}\rangle_\phi\,. \tag{10}$$

It follows that

$$v_{x/\|x\|_p} = \left\|\frac{x}{p}\right\|_p^{1-p}\frac{\tilde{x}}{q}\,. \tag{11}$$

**Proposition 5.2.** *The duality map is a homeomorphism $L_p(M,\phi) \to L_q(M,\phi)$.*

**Proof.** It is immediate from (10) that the duality map is continuous at 0. Further, let $F$ be the map defined in Sec. 2 and $x \neq 0$, then we have from (11)

$$F(x) = \|x\|_p v_{x/\|x\|_p} = \frac{p^p}{pq}\|x\|_p^{2-p}\tilde{x}\,.$$

The statement now follows from Theorem 2.2. □

Let us define the function $\Psi_p : L_p(M,\phi) \to R^+$ by

$$\Psi_p(x) = q\left\|\frac{x}{p}\right\|_p^p = q\varphi(1)\,,$$

where $x = pu\Delta^{1/p}_{\varphi,\phi}$. Then we have

**Proposition 5.3.** *$\Psi_p$ is strongly differentiable. The strong derivative at $x$ is given by*

$$D\Psi_p(x)(y) = \Re\langle y, \tilde{x}\rangle_\phi\,, \qquad y \in L_p(M,\phi)$$

*where $\tilde{x}$ is the dual coordinate. If $1/p + 1/q = 1$, then*

$$\Psi_q(\tilde{x}) = \Re\langle x, \tilde{x}\rangle_\phi - \Psi_p(x)\,.$$

**Proof.** We have from uniform smoothness of $L_p(M,\phi)$ that the norm is strongly differentiable at all points except $x = 0$ and

$$D\|x\|_p(y) = \Re\langle y, v_{x/\|x\|_p}\rangle_\phi\,.$$

It follows from (11) that for $x \neq 0$,

$$D\Psi_p(x)(y) = q\left\|\frac{x}{p}\right\|_p^{p-1} \Re\langle y, v_{x/\|x\|_p}\rangle_\phi = \Re\langle y, \tilde{x}\rangle_\phi\,.$$

As $p > 1$, the function $\|\frac{x}{p}\|_p^p$ is strongly differentiable at $x = 0$ and

$$D\Psi_p(0)(y) = 0 = \Re\langle y, \tilde{0}\rangle_\phi\,.$$

The last equality is rather obvious. □

In the commutative case, as well as on the manifold of positive definite $n \times n$ matrices, $\Psi_p$ is the potential function in the sense of Amari, see Refs. 1 and 14, 11. In general, it is not twice differentiable, but the above proposition shows that the Legendre transformations, relating the dual coordinate systems, are still valid. It will also be clear from the results of the next section, that

$$\Psi_q(\tilde{x}) = \sup_{y\in L_p(M,\phi)} (\Re\langle y, \tilde{x}\rangle_\phi - \Psi_p(y))$$

hence $\Psi_q$ is the conjugate of the convex function $\Psi_p$.

## 6. Divergence in $L_p(M,\phi)$

Following Ref. 1, the function $D_p : L_p(M,\phi) \times L_p(M,\phi) \to R^+$, defined by

$$D_p(x,y) = \Psi_p(x) + \Psi_q(\tilde{y}) - \Re\langle x, \tilde{y}\rangle_\phi$$

is called the divergence. It has the following properties.

**Proposition 6.1.**

(i) *Let $f_p(t) = p + qt^p - pqt$. Then*

$$D_p(x,y) \geq \left\|\frac{y}{p}\right\|_p^p f_p\left(\frac{\|x\|_p}{\|y\|_p}\right) \tag{12}$$

*for all $x$, $y \in L_p(M,\phi)$, where for $y = 0$, we take the limit $\lim_{t\to 0} t^p f_p(s/t) = qs^p$ for all $s$. In particular, $D_p(x,y) \geq 0$ for all $x$, $y \in L_p(M,\phi)$ and equality is attained if and only if $x = y$.*

(ii) *$D_p$ is jointly continuous and strongly differentiable in the first variable.*

(iii) $D_p(y,x) = D_q(\tilde{x},\tilde{y})$.
(iv) $D_p(x,y) + D_p(y,z) = D_p(x,z) + \Re\langle x-y, \tilde{z}-\tilde{y}\rangle_\phi$.

**Proof.** The statement (ii) follows from Propositions 5.2 and 5.3. (iii) and (iv) follow easily from the definition of $D_p$. We will now prove (i). If $y = 0$, then $D_p(x,y) = \Psi_p(x)$ and if $x = 0$, $D_p(x,y) = \Psi_q(\tilde{y})$, which is equal to the right-hand side of (12).

Let now $x \neq 0$, $y \neq 0$ and let $t = \|x\|_p/\|y\|_p$. Then by (11)

$$\Re\langle x,\tilde{y}\rangle_\phi = tq\left\|\frac{y}{p}\right\|_p^{p-1} \Re\left\langle \frac{x}{t}, v_{y/\|y\|_p}\right\rangle_\phi .$$

Let $\|y\|_p = r$ and let $S_r$ be the sphere with radius $r$ in $L_p(M,\phi)$. Then $y$, $\frac{x}{t} \in S_r$. From Sec. 2, the tangent hyperplane $y+H$ to $S_r$ at $y$ is given by $\Re\langle z, v_{y/r}\rangle_\phi = r$, $S_r$ lies entirely in the half-space given by $\Re\langle z, v_{y/r}\rangle_\phi \leq r$ and $y$ is the unique point of $S_r$ contained in $y+H$. Hence,

$$D_p(x,y) \geq \Psi_p(x) + \Psi_q(\tilde{y}) - tpq\left\|\frac{y}{p}\right\|_p^p = \left\|\frac{y}{p}\right\|_p^p f_p(t) \geq 0 ,$$

where equality is attained in the first inequality if and only if $\frac{x}{t} = y$, and in the second inequality if and only if $t = 1$. □

We will also need the following lemma.

**Lemma 6.1.** *Let $y \in L_p(M,\phi)$, $d > 0$ and let*

$$U_{y,d} := \{x \in L_p(M,\phi), D_p(x,y) \leq d\} .$$

*Then $U_{y,d}$ is convex and weakly compact.*

**Proof.** It is easy to see that $D_p$ is convex in the first variable, therefore the set $U_{y,d}$ is also convex. To show that it is weakly compact, it is sufficient to prove that it is weakly closed and norm bounded.

Let $\{x_n\}$ be a sequence in $U_{y,d}$, converging weakly to some $x \in L_p(M,\phi)$. Then $\|x\|_p \leq \liminf_{n\to\infty} \|x_n\|_p$. We therefore have

$$\begin{aligned} D_p(x,y) &= \Psi_q(\tilde{y}) + q\left\|\frac{x}{p}\right\|_p^p - \langle x,\tilde{y}\rangle_\phi \\ &\leq \liminf_{n\to\infty}(\Psi_q(\tilde{y}) + q\left\|\frac{x_n}{p}\right\|_p^p - \langle x_n,\tilde{y}\rangle_\phi) \\ &= \liminf_{n\to\infty} D_p(x_n,y) \leq d \end{aligned}$$

and $U_{y,d}$ is weakly closed.

Finally, for $x \in U_{y,d}$, we have by Proposition 6.1(i),

$$\left\|\frac{y}{p}\right\|_p^p f_p\left(\frac{\|x\|_p}{\|y\|_p}\right) \le D_p(x,y) \le d\,.$$

As $f_p(t)$ goes to infinity for $t \to \infty$, we see that $\|x\|_p$ is bounded. □

## 7. $D_p$-Projections

Let $C$ be a subset in $L_p(M,\phi)$, $y \in L_p(M,\phi)$. If there is a point $x_m \in C$, such that

$$D_p(x_m, y) = \min_{x\in C} D_p(x,y)\,,$$

then $x_m$ will be called a $D_p$-projection of $y$ to $C$. In this section, we prove some uniqueness and existence results for $D_p$-projections.

**Proposition 7.1.** *Let $C$ be a convex subset in $L_p(M,\phi)$, $y \in L_p(M,\phi)$ and $x_m \in C$. The following are equivalent.*

(i) $D_p(x_m, y) = \min_{x\in C} D_p(x,y)$,
(ii) $\tilde{y} - \tilde{x}_m$ *is in the normal cone to $C$ at $x_m$, that is,*

$$\Re\langle x - x_m, \tilde{y} - \tilde{x}_m\rangle_\phi \le 0\,, \qquad \forall\, x \in C$$

(iii) $D_p(x,y) \ge D_p(x,x_m) + D_p(x_m,y)$, $\forall\, x \in C$.

*If such a point exists, it is unique.*

**Proof.** Let $x_m$ be a point in $C$ satisfying (i) and let $x \in C$. Then $x_t = tx + (1-t)x_m$ lies in $C$ for all $t \in [0,1]$ and thus $D_p(x_t, y) \ge D_p(x_m, y)$ on [0, 1]. We have from Proposition 5.3

$$0 \le \frac{d}{dt+} D_p(x_t, y)|_{t=0} = \Re\langle x - x_m, \tilde{x}_m - \tilde{y}\rangle_\phi$$

which is (ii). Further, from Proposition 6.1(iv)

$$\Re\langle x - x_m, \tilde{x}_m - \tilde{y}\rangle_\phi = D_p(x,y) - D_p(x,x_m) - D_p(x_m,y)\,,$$

hence (ii) implies (iii). Finally, let $x_m$ satisfy (iii), then we clearly have $D_p(x_m,y) \le D_p(x,y)$, for all $x \in C$.

To prove uniqueness, suppose that $x_1$ and $x_2$ are points in $C$, satisfying (iii). Then

$$D_p(x_1,y) \ge D_p(x_1,x_2) + D_p(x_2,y) \ge D_p(x_1,x_2) + D_p(x_2,x_1) + D_p(x_1,y)\,.$$

It follows that $D_p(x_1,x_2) + D_p(x_2,x_1) \le 0$ and hence $x_1 = x_2$. □

**Proposition 7.2.** *Let $C \neq \emptyset$ be a weakly closed subset in $L_p(M,\phi)$ and $y \in L_p(M,\phi)$. Then there exists a $D_p$-projection of $y$ to $C$.*

**Proof.** For some $d > 0$, the set $U_{y,d}$ has a nonempty intersection with $C$. By Lemma 6.1, the sets $U_{y,d} \cap C$ are weakly compact. The intersection of these sets for all such $d$ is therefore nonempty and is equal to some $U_{y,\rho} \cap C$. Then $\rho = \min_{x \in C} D_p(x, y)$ and all the points in $U_{y,\rho} \cap C$ are $D_p$-projections of $y$ in $C$. □

By Propositions 7.1 and 7.2, if $C$ is a nonempty convex weakly closed subset in $L_p(M, \phi)$, we can define the map $y \mapsto x_m$, which sends each point $y$ to its unique $D_p$-projection in $C$.

**Proposition 7.3.** *Let $C \neq \emptyset$ be a weakly closed convex subset in $L_p(M, \phi)$. Then the $D_p$-projection is continuous from $L_p(M, \phi)$ with its norm topology to $C$ with the relative weak topology.*

**Proof.** Let $\{y^n\}$ be a sequence in $L_p(M, \phi)$ converging in norm to $y$. Let $x^n_m$ be the unique $D_p$-projection of $y^n$ and $x_m$ be the unique $D_p$-projection of $y$ in $C$, obtained by Propositions 7.2 and 7.1. We have to prove that $x^n_m$ converges weakly to $x_m$.

As the duality map is continuous, we have $\tilde{y}^n \to \tilde{y}$ in $L_q(M, \phi)$. Further, we have from joint continuity of $D_p$ that $\lim D_p(z, y^n) = D_p(z, y)$ and $\lim D_p(y^n, z) = D_p(y, z)$ for any $z \in L_p(M, \phi)$, in particular, $\lim D_p(y^n, y) = \lim D_p(y, y^n) = 0$.

Let $k_1 > 0$ be such that $\|y\| \leq k_1$ and $\|y^n\| \leq k_1$ for all $n$. Further, let us choose an element $z \in C$. By Proposition 6.1(i) and the definition of the $D_p$-projection, we have

$$f_p\left(\frac{\|x^n_m\|}{\|y^n\|}\right) \leq \frac{D_p(z, y^n)}{\|y^n/p\|^p}, \tag{13}$$

$$f_p\left(\frac{\|x_m\|}{\|y\|}\right) \leq \frac{D_p(z, y)}{\|y/p\|^p}. \tag{14}$$

As the right-hand side of (13) converges to the right-hand side of (14), we see that there is a constant $k_2 > 0$, such that $\|x_m\| \leq k_2$ and $\|x^n_m\| \leq k_2$ for all $n$. For sufficiently large $n$, we get by Proposition 6.1(iv)

$$\begin{aligned} d_n := D_p(x^n_m, y^n) &= \inf_{x \in C, \|x\|_p \leq k_2} D_p(x, y^n) \\ &= \inf_{x \in C, \|x\|_p \leq k_2} \{D_p(x, y) + D_p(y, y^n) - \Re\langle x - y, \tilde{y}^n - \tilde{y}\rangle_\phi\} \\ &\leq D_p(x_m, y) + D_p(y, y^n) + (k_1 + k_2)\|\tilde{y} - \tilde{y}^n\|_q \leq d + \varepsilon, \end{aligned}$$

where $d := D_p(x_m, y)$. Similarly,

$$\begin{aligned} D_p(x^n_m, y) &= D_p(x^n_m, y^n) + D_p(y^n, y) - \Re\langle x^n_m - y^n, \tilde{y} - \tilde{y}^n\rangle_\phi \\ &\leq d_n + D_p(y^n, y) + (k_1 + k_2)\|\tilde{y} - \tilde{y}^n\|_q \leq d + 2\varepsilon. \end{aligned}$$

Hence for sufficiently large $n$, $x^n_m \in U_{y,d+2\varepsilon} \cap C$. These sets are nonempty and weakly compact, therefore $\{x^n_m\}$ contains a weakly convergent subsequence. On the other hand, any limit of such subsequence has to be in $U_{y,d+2\varepsilon} \cap C$ for all $\varepsilon$ and thus also in $\bigcap_\varepsilon U_{y,d+2\varepsilon} \cap C$. By uniqueness, this intersection contains a single point $x_m$, it follows that $x_m^n$ converges weakly to $x_m$. □

## 8. The $\alpha$-Divergence in $M_*$

Let $\alpha \in (-1, 1)$ and let $p = \frac{2}{1-\alpha}$. The divergence in $L_p(M, \phi)$ defines the functional $S_\alpha : M_* \times M_* \to R^+$,

$$S_\alpha(\omega_1, \omega_2) := D_p(\ell_\alpha(\omega_1), \ell_\alpha(\omega_2))$$
$$= q\varphi(1) + p\psi(1) - pq\Re\langle u\Delta_{\varphi,\phi}^{1/p}, v\Delta_{\psi,\phi}^{1/q}\rangle_\phi ,$$

where $\omega_1(a) = \varphi(au)$ and $\omega_2(a) = \psi(av)$ are the polar decompositions. It is called the $\alpha$-divergence. Let us remark that a similar definition for the $\alpha$-divergence appeared also in the Discussion in Ref. 19.

It follows from (9) that $S_\alpha$ does not depend of $\phi$. In particular, if $\psi$ is faithful, then

$$\langle u\Delta_{\varphi,\phi}^{1/p}, v\Delta_{\psi,\phi}^{1/q}\rangle_\phi = (\Delta_{\varphi,\xi_\psi}^{1/(2p)}\xi_\psi, \Delta_{\varphi,\xi_\psi}^{1/(2p)}u^*v\xi_\psi) ,$$

where $\xi_\psi$ is a vector representative of $\psi$. It follows that if $\varphi$, $\psi \in M_*^+$, $\psi$ is faithful and $\Delta_{\varphi,\xi_\psi} = \int \lambda E_\lambda$ is the spectral decomposition, then

$$S_\alpha(\varphi, \psi) = (g_p(\Delta_{\varphi,\xi_\psi})\xi_\psi, \xi_\psi) = \int g_p(\lambda)\|E_\lambda \xi_\psi\|^2 ,$$

where $g_p(t) = p + qt - pqt^{1/p}$. Hence, in this case the $\alpha$-divergence is equal to the quasi entropy $S_{g_p}^1$, defined by Petz in Ref. 20 (see also Ref. 18). We will show that this is true on the whole of $M_*^+ \times M_*^+$.

**Lemma 8.1.** *Let $\varphi$, $\psi \in M_*^+$, $u$, $v \in M$ be partial isometries satisfying $u^*u = s(\varphi)$, $v^*v = s(\psi)$. Let $p$, $q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$\langle u\Delta_{\varphi,\phi}^{1/p}, v\Delta_{\psi,\phi}^{1/q}\rangle_\phi = (\Delta_{\varphi,\xi_\psi}^{1/(2p)}\xi_\psi, \Delta_{\varphi,\xi_\psi}^{1/(2p)}u^*v\xi_\psi) , \tag{15}$$

*where $\xi_\psi$ is a vector representative of $\psi$.*

**Proof.** Let $1/p \le 1/2$. We have

$$\langle u\Delta_{\varphi,\phi}^{1/p}, v\Delta_{\psi,\phi}^{1/q}\rangle_\phi = \lim_{y\to 1}(\Delta_{\psi,\phi}^{1/2-1/p}v^*u\Delta_{\varphi,\phi}^{1/p}\eta_\phi(y), \Delta_{\psi,\phi}^{1/2}\eta_\phi(y)) ,$$

with $y \in M_0$, $\|y\| \le 1$. For $y \in N_\phi$,

$$(\Delta_{\psi,\phi}^{1/2-1/p}v^*u\Delta_{\varphi,\phi}^{1/p}\eta_\phi(y), \Delta_{\psi,\phi}^{1/2}\eta_\phi(y))$$
$$= (J_{\xi_\psi,\eta_\phi}\Delta_{\psi,\phi}^{1/2}\eta_\phi(y), J_{\xi_\psi,\eta_\phi}\Delta_{\psi,\phi}^{1/2-1/p}v^*u\Delta_{\varphi,\phi}^{1/p}\eta_\phi(y))$$
$$= (y^*\xi_\psi, J_{\xi_\psi,\eta_\phi}\Delta_{\psi,\phi}^{1/2-1/p}v^*u\Delta_{\varphi,\phi}^{1/p}\eta_\phi(y)) , \tag{16}$$

here we have used that $J^*_{\xi_\psi,\eta_\phi} J_{\xi_\psi,\eta_\phi} = s(\psi) = s(\Delta_{\psi,\phi})$, the support of $\Delta_{\psi,\phi}$. Let $t \in \mathbb{R}$, then

$$\begin{aligned} J_{\xi_\psi,\eta_\phi} \Delta_{\psi,\phi}^{1/2-it} v^* u \Delta_{\varphi,\phi}^{it} \eta_\phi(y) &= S_{\xi_\psi,\eta_\phi} \Delta_{\psi,\phi}^{-it} v^* u \Delta_{\varphi,\phi}^{it} \eta_\phi(y) \\ &= S_{\xi_\psi,\eta_\phi} v^* \Delta_{\psi_v,\phi}^{-it} \Delta_{\varphi_u,\phi}^{it} u \eta_\phi(y) \\ &= S_{\xi_\psi,\eta_\phi} v^* (D\psi_v : D\varphi_u)_{-t} u \eta_\phi(y) \\ &= y^* u^* (D\psi_v : D\varphi_u)^*_{-t} v \xi_\psi \,, \end{aligned}$$

where $\varphi_u(a) = \varphi(u^* a u)$ and $u \Delta_{\varphi,\phi}^{it} u^* = \Delta_{\varphi_u,\phi}^{it}$ by (C.8) in Ref. 4. From this, we have

$$\begin{aligned} (y^* \xi_\psi, J_{\xi_\psi,\eta_\phi} \Delta_{\psi,\phi}^{1/2-it} v^* u \Delta_{\varphi,\phi}^{it} \eta_\phi(y)) &= ((D\psi_v : D\varphi_u)_{-t} u y y^* \xi_\psi, v \xi_\psi) \\ &= (\Delta_{\psi_v,\xi_\psi}^{-it} \Delta_{\varphi_u,\xi_\psi}^{it} u y y^* \xi_\psi, v \xi_\psi) \\ &= (u \Delta_{\varphi,\xi_\psi}^{it} y y^* \xi_\psi, v \xi_\psi) \,, \end{aligned}$$

where we have used (C.5) and (C.8) of Ref. 4. It follows that for $z = it$,

$$(y^* \xi_\psi, J_{\xi_\psi,\eta_\phi} \Delta_{\psi,\phi}^{1/2-z} v^* u \Delta_{\varphi,\phi}^{z} \eta_\phi(y)) = (y^* \xi_\psi, y^* \Delta_{\varphi,\xi_\psi}^{\bar{z}} u^* v \xi_\psi) \,. \tag{17}$$

By Lemma 3.1 in Ref. 16, both sides of (17) are holomorphic for $0 < \Re z < 1/2$ and continuous for $0 \leq \Re z \leq 1/2$. Equation (15) holds for $1/p \leq 1/2$ by (16) and analytic continuation of (17).

Let now $1/q \leq 1/2$. We have by the first part of the proof

$$\begin{aligned} \langle u \Delta_{\varphi,\phi}^{1/p}, v \Delta_{\psi,\phi}^{1/q} \rangle_\phi &= (u \xi_\varphi, v \Delta_{\xi_\psi,\xi_\varphi}^{1/q} \xi_\varphi) = (S_{\xi_\varphi,\xi_\psi} u^* v \xi_\psi, \Delta_{\xi_\psi,\xi_\varphi}^{1/q} S_{\xi_\varphi,\xi_\psi} \xi_\psi) \\ &= (J_{\xi_\psi,\xi_\varphi} \Delta_{\xi_\psi,\xi_\varphi}^{1/q} J_{\xi_\varphi,\xi_\psi} \Delta_{\xi_\varphi,\xi_\psi}^{1/2} \xi_\psi, \Delta_{\xi_\varphi,\xi_\psi}^{1/2} u^* v \xi_\psi) \\ &= (\Delta_{\varphi,\xi_\psi}^{1/p-1/2} \xi_\psi, \Delta_{\varphi,\xi_\psi}^{1/2} u^* v \xi_\psi) \,, \end{aligned}$$

we have used the equations (C.14) $J^*_{\eta_1,\eta_2} = J_{\eta_2,\eta_1}$ and ($\beta$5) $J_{\eta_1,\eta_2} \Delta_{\eta_1,\eta_2} J_{\eta_2,\eta_1} = \Delta^{-1}_{\eta_2,\eta_1}$ from Appendix C in Ref. 4. □

It follows that $S_\alpha(\varphi,\psi) = S^1_{g_p}(\varphi,\psi)$ for all positive normal functionals $\varphi$ and $\psi$. The function $g_p$, $1 < p < \infty$ is operator convex and it follows from the results in Ref. 20 that

(i) $S_\alpha$ is jointly convex on $M_*^+ \times M_*^+$,
(ii) $S_\alpha$ decreases under stochastic maps on $M_*^+ \times M_*^+$,
(iii) $S_\alpha$ is lower semicontinuous on $M_*^+ \times \mathcal{F}(M_*^+)$ endowed with the product of norm topologies, where $\mathcal{F}(M_*^+)$ denotes the set of faithful elements in $M_*^+$.

The following properties of the $\alpha$-divergence are valid on $M_* \times M_*$ and are immediate consequences of the results of Sec. 6.

(i) Positivity

$$S_\alpha(\varphi,\psi) \geq \|\psi\|_1 g_p\left(\frac{\|\varphi\|_1}{\|\psi\|_1}\right) \geq 0$$

and $S_\alpha(\varphi,\psi) = 0$ if and only if $\varphi = \psi$ (here $\|\cdot\|_1$ is the norm in $M_*$).

(ii) $S_\alpha(\varphi,\psi) = S_{-\alpha}(\psi,\varphi)$

(iii) generalized Pythagorean relation

$$S_\alpha(\varphi,\psi) + S_\alpha(\psi,\sigma) = S_\alpha(\varphi,\sigma) + \Re\langle \ell_\alpha(\varphi) - \ell_\alpha(\psi), \ell_{-\alpha}(\sigma) - \ell_{-\alpha}(\psi)\rangle_\phi\,.$$

Let $\varphi$, $\psi \in M_*$ and let $x_t$, $t \in [0,1]$ be a curve in $L_p(M,\phi)$, given by

$$x_t := \ell_\alpha(\varphi) + t(\ell_\alpha(\psi) - \ell_\alpha(\varphi))\,,$$

then $\ell_\alpha^{-1}(x_t)$ is the $\alpha$-geodesic in $M_*$, connecting $\varphi$ and $\psi$. Notice that the Pythagorean relation (iii) is a generalization of the classical version in Ref. 1, which says that equality is attained if and only if the $\alpha$-geodesic connecting $\psi$ and $\varphi$ is orthogonal to the $-\alpha$-geodesic connecting $\psi$ and $\sigma$.

We also define the $\alpha$-projection of $\varphi \in M_*$ onto a subset $C \subset M_*$ as the element in $C$ that minimizes $S_\alpha(\cdot,\varphi)$ over $C$. We will say that a subset $C \subset M_*$ is $\alpha$-convex if $\ell_\alpha(C)$ is convex. The next proposition is a generalization of the results in Refs. 1 and 2 and follows directly from Proposition 7.1.

**Proposition 8.1.** *Let $C \subset M_*$ be $\alpha$-convex and let $\psi \in M_*$, $\varphi_m \in C$. The following are equivalent:*

(i) *$\varphi_m$ is an $\alpha$-projection of $\psi$ in $C$.*

(ii) *For all $\sigma \in C$,*

$$S_\alpha(\sigma,\psi) \geq S_\alpha(\varphi_m,\psi) + S_{-\alpha}(\varphi_m,\sigma)\,.$$

(iii) *If $\psi_t$ is the $-\alpha$-geodesic connecting $\varphi_m$ and $\psi$, then $\frac{d}{dt}\ell_{-\alpha}(\psi_t)$ lies in the normal cone to $\ell_\alpha(C)$ at $\ell_\alpha(\varphi_m)$.*

*If such a point exists, it is unique.*

The topology induced by the $\alpha$-embedding from the norm, (resp. the weak topology in $L_p(M,\phi)$) will be called the $\alpha$-, (resp. the $\alpha$-weak topology). The following proposition is also immediate from Sec. 7.

**Proposition 8.2.** *Let $C$ be a nonempty subset in $M_*$ and let $\psi \in M_*$.*

(i) *If $C$ is $\alpha$-weakly closed, then there exists an $\alpha$-projection of $\psi$ in $C$.*

(ii) *If $C$ is $\alpha$-convex, $\alpha$-weakly closed, then the $\alpha$-projection is a continuous map from $M_*$ with the $\alpha$-topology to $C$ with the relative $\alpha$-weak topology.*

## 9. The Case $\alpha = 0$

Let $\alpha = 0$, $p = q = 2$. The space $L_2(M, \phi)$ can be identified with the Hilbert space $H_\phi$ and the dual pairing $\langle \cdot, \cdot \rangle_\phi$ is the inner product $(\cdot, \cdot)$ in $H_\phi$. Through this identification, the 0-embedding becomes the map

$$\omega \mapsto 2u\xi_\varphi ,$$

where $\omega(a) = \varphi(au)$ is the polar decomposition of $\omega$ and $\xi_\varphi$ is the unique vector representative of $\varphi$ in the natural positive cone $V$ in $H_\phi$. Hence the 0-embedding maps $M_*$ bijectively onto $H_\phi$. Up to multiplication by 2, the restriction of $\ell_0$ to the positive cone $M_*^+$ corresponds to the identification of the positive normal functionals with elements in $V$ proved by Araki in Ref. 3. It has also been shown that this identification is a homeomorphism $M_*^+ \to V$. It follows that the relative 0-topology is the same as the relative $L_1$-topology in $M_*^+$.

The duality map is the identity on $H_\phi$ and the potential function is

$$\Psi_2(x) = \frac{1}{2}\|x\|^2 .$$

Therefore, the potential function is $C^\infty$-diferentiable and

$$D^2\Psi_2(x)(y, z) = \Re(y, z) \qquad \forall\, x \in H_\phi .$$

It follows that $\Psi_2$ defines a Riemannian metric in the tangent bundle $T\mathcal{M}_0$, which corresponds to the real part of the inner product, induced from the 0-embedding. In the matrix case, this metric was studied on density matrices and it was shown that it coincides with the Wigner–Yanase metric, see Ref. 10.

The $D_2$-divergence in $H_\phi$ is

$$D_2(x, y) = \frac{1}{2}\|x - y\|^2 ,$$

hence the $D_2$-projection corresponds to minimizing the Hilbert space norm. This means, in particular, that there is a unique $D_2$-projection onto every closed convex subset of $H_\phi$.

The 0-divergence in $M_*$ becomes

$$S_0(\omega_1, \omega_2) = 2\|u\xi_\varphi - v\xi_\psi\|^2 .$$

On the positive cone, the 0-divergence generalizes the classical Hellinger distance.

## 10. The Unit Sphere

The $\alpha$-embedding maps the unit sphere $S$ in $M_*$ onto the sphere $S_p$ with radius $p$ in $L_p(M, \phi)$. The duality map $x \mapsto \tilde{x}$ maps $S_p$ onto the sphere $S'_q$ with radius $q$ in the dual space $L_q(M, \phi)$. From (11), we have that for $x \in S_p$,

$$\tilde{x} = qv_{x/p} . \tag{18}$$

**Proposition 10.1.** *The duality map $S_p \ni x \mapsto \tilde{x} \in S'_q$ is uniformly continuous.*

**Proof.** The statement follows from (18) and Theorem 2.1. $\square$

For each $x \in S_p$, there is a unique tangent hyperplane $x + H_x$ through $x$, where $H_x$ is given by the condition

$$\Re\langle y, \tilde{x}\rangle_\phi = q\Re\langle y, v_{x/p}\rangle_\phi = 0\,.$$

Hence there is a splitting $L_p(M,\phi) = H_x \oplus [x]$ and, similarly as in Ref. 8, there is a continuous projection $\pi_x : L_p(M,\phi) \to H_x$, given by

$$\pi_x(y) = y - \Re\langle y, v_{x/\|x\|_p}\rangle_\phi \frac{x}{p} = y - \frac{1}{pq}\Re\langle y, \tilde{x}\rangle_\phi x\,,$$

which is obtained by minimizing the $L_p$-norm.

As the norm is strongly differentiable, the unit sphere forms a differentiable submanifold $\mathcal{D}_\alpha$ in $\mathcal{M}_\alpha$. Let $\omega \in \mathcal{D}_\alpha$ and let $x \in S_p$ be its $\alpha$-coordinate. The tangent space $T_x(\mathcal{D}_\alpha)$ can be identified with the tangent hyperplane $H_x$ and $\pi_x$ can be used to project the $\alpha$-connection onto $T\mathcal{D}_\alpha$.

**Remark 10.1.** Let $\psi$ be a faithful state with the $\alpha$-coordinate $x$ in $L_p(M,\phi)$ and let $\xi_\psi$ be the vector representative of $\psi$ in a natural positive cone. Then $\psi$ has the $\alpha$-coordinate $\tau_p(\psi,\phi)(x) = p\Delta_{\xi_\psi}^{1/p}$ in $L_p(M,\psi)$ and the dual coordinate $\tau_q(\psi,\phi)(\tilde{x}) = q\Delta_{\xi_\psi}^{1/q}$ in $L_q(M,\psi)$. By (9) we have for $y \in L_p(M,\phi)$,

$$\langle y, \tilde{x}\rangle_\phi = \langle \tau_p(\psi,\phi)(y), q\Delta_{\xi_\psi}^{1/q}\rangle_\psi = (y_\psi\xi_\psi, q\Delta_{\xi_\psi}^{1/q}\xi_\psi) = q\langle y_\psi, 1\rangle_\psi\,,$$

where $y_\psi = \tau_p(\psi,\phi)(y)$. It follows that $\tau_p(\psi,\phi)$ maps $T_x(\mathcal{D}_\alpha)$ onto the subspace

$$\mathcal{F}_\psi^\alpha := \{z \in L_p(M,\psi), \Re\langle z, 1\rangle_\psi = 0\}\,.$$

This corresponds to the results in Ref. 9, where the $\alpha$-connection is obtained on the fiber bundle $\mathcal{F}^\alpha$ over a manifold of faithful states.

The projected connection, even in the classical and the matrix case, is no longer flat. Hence, it does not define a divergence, but nevertheless, we can use the restriction of $S_\alpha$ as a quasi-distance on the unit sphere. This restriction has the form

$$S_\alpha(\omega_1,\omega_2) = pq(1 - \Re\langle u\Delta_{\varphi,\phi}^{1/p}, v\Delta_{\psi,\phi}^{1/q}\rangle_\phi)$$

which corresponds to the definition of the $\alpha$-divergence in Ref. 1 for probability densities and in Ref. 13 for density matrices.

## Acknowledgments

The paper was done during the stay in RIKEN Brain Science Institute, Tokyo. The author wishes to thank Professor S. Amari for providing the financial support, H. Hasegawa for arranging the visit and I. Ojima and D. Petz for valuable comments. The research was supported by the grant VEGA 1/0264/03 and Science and Technology Assistance Agency under the contract No. APVT-51-032002.

## References


1. S. Amari, *Differential-Geometrical Methods in Statistic*, Lecture Notes in Statistics, Vol. 28 (1985).
2. S. Amari and H. Nagaoka, *Method of Information Geometry*, AMS Monograph (Oxford Univ. Press, 2000).
3. H. Araki, Some properties of modular conjugation operator of von Neumann algebras and a non-commutative Radon–Nikodym theorem with a chain rule, *Pac. J. Math.* **50** (1974) 309–354.
4. H. Araki and T. Masuda, Positive cones and $L_p$-spaces for von Neumann algebras, *Publ. RIMS, Kyoto Univ.* **18** (1982) 339–411.
5. N. N. Chentsov, *Statistical Decision Rules and Optimal Inferences*, Trans. Math. Monog., Vol. 53 (Amer. Math. Soc., 1982).
6. D. F. Cudia, The geometry of Banach spaces. Smoothness, *Trans. Amer. Math. Soc.* **110** (1964) 284–314.
7. J. Diestel, *Geometry of Banach Spaces — Selected Topics*, Lecture Notes in Math., Vol. 485 (Springer, 1975).
8. P. Gibilisco and G. Pistone, Connections on non-parametric statistical manifolds by Orlicz space geometry, *Inf. Dim. Anal. Quantum Probab. Rel. Topics* **1** (1998) 325–347.
9. P. Gibilisco and T. Isola, Connections on statistical manifolds of density operators by geometry of non-commutative $L_p$-spaces, *Inf. Dim. Anal. Quantum Probab. Rel. Topics* **2** (1999) 169–178.
10. P. Gibilisco and T. Isola, A characterization of Wigner Yanase skew information among statistically monotone metrics, *Inf. Dim. Anal. Quantum Probab. Rel. Topics* **4** (2001) 553-557.
11. M. R. Grasselli, Monotonicity, Duality and uniqueness of the WYD metrics, *Inf. Dim. Anal. Quantum Probab.* **7** (2004) 215–232.
12. M. Grasselli and R. F. Streater, The quantum information manifold for epsilon-bounded forms, *Rep. Math. Phys.* **46** (2000) 325–335.
13. H. Hasegawa, $\alpha$-divergence of the non-commutative information geometry, *Rep. Math. Phys.* **33** (1993) 87–93.
14. A. Jenčová, Geometry of quantum states: dual connections and divergence functions, *Rep. Math. Phys.* **47** (2001) 121–138.
15. G. Köthe, *Topological Vector Spaces 1* (Springer-Verlag, 1983).
16. T. Masuda, $L_p$-spaces for von Neumann algebra with reference to a faithful normal semifinite weight, *Publ. RIMS, Kyoto Univ.* **19** (1983) 673–727.
17. H. Nagaoka, Differential geometrical aspects of quantum state estimation and relative entropy, in *Quantum Communication, Computing and Measurement*, eds. O. Hirota *et al.* (Plenum Press, 1994).
18. M. Ohya and D. Petz, *Quantum Entropy and Its Use* (Springer, 1993).
19. I. Ojima, Temperature as order parameter of broken scale invariance, to appear in *Publ. RIMS, Kyoto Univ.*
20. D. Petz, Quasi-entropies for states of a von Neumann algebra, *Publ. RIMS, Kyoto Univ.* **21** (1985) 787–800.
21. D. Petz, Information Geometry of quantum states, in: *Quantum Probab. Commun.*, Vol. 10, eds. R. L. Hudson *et al.* (World Scientific, 1998), pp. 135–158.
22. G. Pistone and C. Sempi, An infinite dimensional geometric structure on the space of all probability measures equivalent to a given one, *Ann. Statist.* **23** (1995) 1543–1561.

23. R. F. Streater, The information manifold for relatively bounded potentials, *Tr. Mat. Inst. Stek.* **228** (2000) 217–235.
24. R. F. Streater, The analytic quantum information manifold, in: *Stochastic Processes, Physics and Geometry: New Interplays, II (Leipzig 1999)* (Amer. Math. Soc., 2000), pp. 603–611.

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Functional Analysis 239 (2006) 1–20
JOURNAL OF Functional Analysis
www.elsevier.com/locate/jfa

# A construction of a nonparametric quantum information manifold

Anna Jenčová [1]

*Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, 814 73 Bratislava, Slovakia*

Received 9 December 2005; accepted 16 February 2006

Available online 23 March 2006

Communicated by D. Voiculescu

**Abstract**

We present a construction of a Banach manifold structure on the set of faithful normal states of a von Neumann algebra, where the underlying Banach space is a quantum analogue of an Orlicz space. On the manifold, we introduce the exponential and mixture connections as dual pair of affine connections.


*Keywords:* Information manifold; Dual connections; Relative entropy; State perturbation; Quantum Orlicz space

## 1. Introduction

An information manifold is a family of states of some classical or quantum system, endowed with a differentiable manifold structure. For parametrized families of probability distributions, the geometry of such manifolds and its applications in parameter estimation is already well understood, see, for example, the books [3,4]. This development was started by Rao [20] and Jeffreys [11], who suggested the Fisher information as a Riemannian metric for parametrized statistical models. Later on, Effron [5] defined the concept of statistical curvature and pointed out the importance of exponential models, which led to the introduction of the exponential affine connection on the manifold. Amari in his well-known book [2] equipped the manifold with a family of $\alpha$-connections and introduced the concept of duality, which is related to the notion

*E-mail address:* jenca@mat.savba.sk.

[1] The research was supported by the grant VEGA No. 1/3016/06, Center of Excellence SAS Physics of Information I/2/2005 and Science and Technology Assistance Agency under the contract No. APVT-51-032002.

0022-1236/$ – see front matter 
doi:10.1016/j.jfa.2006.02.007

of statistical divergence. The most important of these structures, in applications as well as in the theory, is the dual pair of exponential and mixture connections, with the related statistical divergence, called the I-divergence, or relative entropy.

The nonparametric information manifold was introduced by Pistone and Sempi [18,19], based on the idea of a nonparametric exponential model. As it turned out, a natural parametrization for such models is given by the exponential Orlicz space. Further developments, including the definition of the affine connections and duality, can be found in [7,8].

For families of quantum states, similar structures were found in the finite-dimensional case, see, for example, [10,12,15,17]. In infinite dimensions, the situation is more complicated. As there is no suitable noncommutative counterpart of the exponential Orlicz space, it is not clear how to choose the underlying Banach space for the manifold. Some suggestions can be found in [9,13,21,22]. See also [1] for a definition of a noncommutative Orlicz space.

The aim of this paper is to introduce a differentiable manifold structure on the set of faithful states of a quantum system, represented by a von Neumann algebra $\mathcal{M}$. Moreover, we want this manifold to be a quantum counterpart of the Pistone and Sempi construction.

We use an approach similar to Grasselli [8] in the commutative case: we define an Orlicz norm on the space of self-adjoint operators in $\mathcal{M}$ and take the completion under this norm to be the underlying Banach space for the manifold. The norm is defined by a quantum Young function, as in [23]. The definition of a Young function on a Banach space, together with some results on the associated norms, can be found in Section 3. For a faithful state $\varphi$, the quantum Orlicz space $B_\varphi$ and its centered version $B_{\varphi,0}$ are introduced in Section 4. The definition of the related Young function is based on the relative entropy approach to state perturbation. We treat the dual spaces in Section 6. The main result is contained in Section 8, where the manifold structure is introduced and, moreover, the exponential and mixture connections are defined as a pair of dual affine connections on each connected component of the manifold.

## 2. Preliminaries

We recall some properties of relative entropy and perturbed states, that will be needed later. See [16] for details.

Let $\mathcal{M}$ be a von Neumann algebra in standard form. For $\omega$ and $\varphi$ in $\mathcal{M}_*^+$, the relative entropy is defined as

$$S(\omega,\varphi)=\begin{cases}-\langle\log(\Delta_{\varphi,\xi_\omega})\xi_\omega,\xi_\omega\rangle & \text{if } \operatorname{supp}\omega\leqslant\operatorname{supp}\varphi,\\ \infty & \text{otherwise,}\end{cases}$$

where $\xi_\omega$ is the representing vector of $\omega$ in a natural positive cone and $\Delta_{\varphi,\xi_\omega}$ is the relative modular operator. Then $S$ is jointly convex and weakly lower semicontinuous. Let us denote $\mathcal{P}_\varphi:=\{\omega\in\mathcal{M}_*^+, S(\omega,\varphi)<\infty\}$, then $\mathcal{P}_\varphi$ is a convex cone. We will need the following identity:

$$S(\psi,\varphi)+\sum_i S(\psi_i,\psi)=\sum_i S(\psi_i,\varphi), \tag{1}$$

where $\psi_i\in\mathcal{M}_*^+$, $i=1,\dots,n$, and $\psi=\sum_i\psi_i$. Since $S(\psi_i,\psi)$ is always finite, it follows from this identity that $\sum_i\psi_i\in\mathcal{P}_\varphi$ if and only if $\psi_i\in\mathcal{P}_\varphi$ for all $i$.

Let $\mathfrak{S}_*$ be the set of normal states on $\mathcal{M}$ and let $\mathcal{S}_\varphi := \{\omega \in \mathfrak{S}_*,\ S(\omega, \varphi) < \infty\}$. Then $\mathcal{S}_\varphi$ is a convex set and generates $\mathcal{P}_\varphi$. From (1), we get

$$S(\psi_\lambda, \varphi) + \lambda S(\psi_1, \psi_\lambda) + (1-\lambda) S(\psi_2, \psi_\lambda) = \lambda S(\psi_1, \varphi) + (1-\lambda) S(\psi_2, \varphi), \tag{2}$$

where $\psi_1$, $\psi_2$ are normal states and $\psi_\lambda = \lambda\psi_1 + (1-\lambda)\psi_2$, $0 \leqslant \lambda \leqslant 1$. As above, it follows that $\psi_\lambda \in \mathcal{S}_\varphi$ if and only if both $\psi_1, \psi_2 \in \mathcal{S}_\varphi$, in other words, $\mathcal{S}_\varphi$ is a face in $\mathfrak{S}_*$. For $C > 0$, we define the set $\mathcal{S}_{\varphi,C} := \{\omega,\ S(\omega, \varphi) \leqslant C\}$. Then $\mathcal{S}_{\varphi,C}$ is convex and compact in the $\sigma(\mathcal{M}_*, \mathcal{M})$ topology.

Let us suppose that $\varphi$ is a faithful normal state on $\mathcal{M}$ and let $h$ be a self-adjoint element in $\mathcal{M}$. The perturbed state $[\varphi^h]$ is defined as the unique maximizer of

$$\sup_{\omega \in \mathfrak{S}_*} \bigl\{\omega(h) - S(\omega, \varphi)\bigr\}. \tag{3}$$

Then $[\varphi^h]$ is a faithful normal state and $S([\varphi^h], \varphi)$ is finite. Let $c_\varphi(h)$ be the supremum in (3), that is

$$c_\varphi(h) = \bigl[\varphi^h\bigr](h) - S\bigl(\bigl[\varphi^h\bigr], \varphi\bigr). \tag{4}$$

It is known that

$$\varphi(h) \leqslant c_\varphi(h) \leqslant \log \varphi\bigl(e^h\bigr). \tag{5}$$

Moreover, we have

$$\omega(h) - S(\omega, \varphi) = c_\varphi(h) - S\bigl(\omega, \bigl[\varphi^h\bigr]\bigr) \tag{6}$$

for any self-adjoint $h \in \mathcal{M}$ and $\omega \in \mathfrak{S}_*$. Let $h, k$ be self-adjoint elements in $\mathcal{M}$, then the chain rule $[\varphi^{h+k}] = [[\varphi^h]^k]$ and

$$c_\varphi(h+k) = c_{[\varphi^h]}(k) + c_\varphi(h) \tag{7}$$

holds. Let now $\xi_\varphi$ be the vector representative of $\varphi$ and let $\varphi^h \in \mathcal{M}_*^+$ be the functional induced by the perturbed vector

$$\xi_\varphi^h := e^{\frac{1}{2}(\log \Delta_\varphi + h)} \xi_\varphi = e^{c_\varphi(h)} \Delta_{[\varphi^h],\varphi}^{\frac{1}{2}} \xi_\varphi.$$

Then $c_\varphi(h) = \log \varphi^h(1)$ and $[\varphi^h] = \varphi^h / \varphi^h(1)$. Moreover, if $\varphi^h = \varphi^k$, then $h = k$.

## 3. Young functions on Banach spaces and the associated norms

Let $V$ be a real Banach space and let $V^*$ be its dual, with the duality pairing $\langle v, x\rangle = v(x)$. Recall that any convex lower semicontinuous function $V \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous with respect to the $\sigma(V, V^*)$-topology.

### 3.1. The Young function

We will say that a function $\Phi : V \to \mathbb{R} \cup \{\infty\}$ is a Young function, if it satisfies:

(i) $\Phi$ is convex and lower semicontinuous,
(ii) $\Phi(x) \geqslant 0$ for all $x \in V$ and $\Phi(0) = 0$,
(iii) $\Phi(x) = \Phi(-x)$ for all $x \in V$,
(iv) if $x \neq 0$, then $\lim_{t\to\infty} \Phi(tx) = \infty$.

**Lemma 3.1.** *Let $\Phi$ be a Young function. Let us define the sets*

$$C_\Phi := \{x \in V, \Phi(x) \leqslant 1\},$$
$$L_\Phi := \{x \in V, \exists s > 0, \text{ such that } \Phi(sx) < \infty\}.$$

*Then $C_\Phi$ is absolutely convex and $L_\Phi = \bigcup_n nC_\Phi$. In particular, $L_\Phi$ is a (real) vector space.*

**Proof.** Let $x, y \in C_\Phi$ and let $\alpha, \beta \in \mathbb{R}$, such that $|\alpha| + |\beta| \leqslant 1$. Put $\gamma = 1 - |\alpha| - |\beta|$, then

$$\Phi(\alpha x + \beta y) = \Phi\big(|\alpha| \operatorname{sgn}(\alpha)x + |\beta| \operatorname{sgn}(\beta)y + \gamma 0\big) \leqslant |\alpha|\Phi(x) + |\beta|\Phi(y) \leqslant 1$$

hence $\alpha x + \beta y \in C_\Phi$ and $C_\Phi$ is absolutely convex.

Let now $x \in L_\Phi$ and let $s > 0$ be such that $\Phi(sx) = K < \infty$. Choose $m \in \mathbb{N}$ such that $m \geqslant \max\{1/s, K/s\}$, then by convexity

$$\Phi\left(\frac{1}{m}x\right) = \Phi\left(\frac{1}{ms}sx\right) \leqslant \frac{1}{ms}\Phi(sx) = \frac{K}{ms} \leqslant 1$$

and $x \in mC_\Phi$. Since obviously $nC_\Phi \subset L_\Phi$ for all $n$, we have $L_\Phi = \bigcup_n nC_\Phi$, which clearly implies that $L_\Phi$ is a vector space. □

Let us recall that the effective domain

$$\operatorname{dom}(\Phi) = \{x \in V, \ \Phi(x) < \infty\}$$

is a convex set. Any convex lower semicontinuous function is continuous in the interior of its effective domain [6]. Clearly, $L_\Phi$ is the smallest vector space containing $\operatorname{dom}(\Phi)$.

In the space $L_\Phi$, we now introduce the Minkowski functional of $C_\Phi$,

$$\|x\|_\Phi := \inf\{\rho > 0, \ x \in \rho C_\Phi\}.$$

Since $C_\Phi$ is absolutely convex and absorbing, $\|\cdot\|_\Phi$ is a seminorm. Moreover, $\|x\|_\Phi = 0$ means that $\Phi(tx) \leqslant 1$ for all $t > 0$. By the property (iv), this implies that $x = 0$. It follows that $\|\cdot\|_\Phi$ defines a norm in $L_\Phi$. Let us denote by $B_\Phi$ the completion of $L_\Phi$ under this norm.

**Lemma 3.2.** *Let $x \in L_\Phi$. Then $\|x\|_\Phi \leqslant 1$ if and only if $\Phi(x) \leqslant 1$.*

**Proof.** If $\|x\|_\Phi < 1$, then $x \in C_\Phi$ and $\Phi(x) \leqslant 1$. Let now $\|x\|_\Phi = 1$ and let $t_n < 1$ be a sequence converging to 1. Then $\Phi(t_n x) \leqslant 1$ for all $n$ and, by lower semicontinuity, $\Phi(x) \leqslant \liminf_n \Phi(t_n x) \leqslant 1$. Hence $\|x\|_\Phi \leqslant 1$ implies $\Phi(x) \leqslant 1$. On the other hand, if $\Phi(x) \leqslant 1$, then $x \in C_\Phi$ and clearly $\|x\|_\Phi \leqslant 1$. $\square$

**Lemma 3.3.** *Let $x \in L_\Phi$. Then $\|x\|_\Phi \leqslant 1$ implies $\Phi(x) \leqslant \|x\|_\Phi$ and $\|x\|_\Phi > 1$ implies $\Phi(x) \geqslant \|x\|_\Phi$. Moreover, if $\Phi$ is finite valued, then $\|x\|_\Phi = 1$ if and only if $\Phi(x) = 1$.*

**Proof.** Let $\|x\|_\Phi \leqslant 1$. By convexity of $\Phi$ and Lemma 3.2,

$$\Phi(x) = \Phi\left(\|x\|_\Phi \frac{x}{\|x\|_\Phi}\right) \leqslant \|x\|_\Phi \Phi\left(\frac{x}{\|x\|_\Phi}\right) \leqslant \|x\|_\Phi.$$

Let now $\|x\|_\Phi > 1$, then $\Phi(x) > 1$. If $\Phi(x) = \infty$, then the assertion is obviously true. Let us suppose that $\Phi(x)$ is finite. The function $t \mapsto \Phi(tx)$ is convex and bounded on $\langle 0, 1\rangle$, hence continuous on $(0, 1)$. It follows that $\Phi(tx) = 1$ for some $t$ in this interval and clearly $t = 1/\|x\|_\Phi$. We have

$$1 = \Phi(tx) \leqslant t\Phi(x)$$

and hence $\|x\|_\Phi \leqslant \Phi(x)$. This also proves that last statement. $\square$

*3.2. The conjugate function*

Let $V^*$ be the dual space. Let the function $\Phi^* : V^* \to \mathbb{R} \cup \{\infty\}$ be the conjugate of $\Phi$,

$$\Phi^*(v) = \sup_{x \in V}\bigl\{v(x) - \Phi(x)\bigr\} = \sup_{x \in \mathrm{Dom}(\Phi)}\bigl\{v(x) - \Phi(x)\bigr\}.$$

The function $\Phi^*$ is convex, lower semicontinuous and positive, $\Phi^*(v) = \Phi^*(-v)$ and $\Phi^*(0) = 0$. But, in general, $\Phi^*$ is not a Young function: consider the case when $\Phi(0) = 0$ and $\Phi(x) = \infty$ for all $x \neq 0$, then $\Phi$ is a Young function, but its conjugate is identically equal to 0 on $V^*$ and the condition (iv) is not satisfied.

Let $(\mathrm{dom}(\Phi))^\perp$ be the orthogonal subspace to $\mathrm{dom}(\Phi)$ in $V^*$, that is

$$\bigl(\mathrm{dom}(\Phi)\bigr)^\perp := \bigl\{v \in V^*,\ v(x) = 0 \text{ for all } x \in \mathrm{dom}(\Phi)\bigr\}.$$

Then $(\mathrm{dom}(\Phi))^\perp$ is a closed subspace in $V^*$. Let $\tilde{V}$ be the quotient space $\tilde{V} = V^*/_{(\mathrm{dom}(\Phi))^\perp}$. If $u$ and $v$ are elements in the same equivalence class, then

$$\Phi^*(v) = \sup_{x \in \mathrm{dom}(\Phi)}\bigl\{v(x) - \Phi(x)\bigr\} = \sup_{x \in \mathrm{dom}(\Phi)}\bigl\{u(x) - \Phi(x)\bigr\} = \Phi^*(u)$$

and $\Phi^*$ is well defined as a function on $\tilde{V}$.

**Lemma 3.4.** $\Phi^* : \tilde{V} \to \mathbb{R} \cup \{\infty\}$ *is a Young function.*

**Proof.** It is easy to see that $\Phi^*$ satisfies (i)–(iii) from the definition of a Young function. Moreover, it follows from the definition of the conjugate function that

$$\left|v(x)\right| \leqslant \Phi(x) + \Phi^*(v) \quad \text{for all } x \in V, v \in \tilde{V}. \tag{8}$$

Let $v \in \tilde{V}$, $v \neq 0$. Then there is an element $x \in \mathrm{dom}(\Phi)$ such that $v(x) \neq 0$. It follows that $\Phi^*(tv) \geqslant |tv(x)| - \Phi(x)$ for all $t$ and (iv) is satisfied. □

We will define $C_{\Phi^*}$, $L_{\Phi^*}$, $\|\cdot\|_{\Phi^*}$ and $B_{\Phi^*}$ in the same way as for $\Phi$.

**Lemma 3.5** *(Hölder inequality).*

$$\left|v(x)\right| \leqslant 2\|x\|_\Phi \|v\|_{\Phi^*} \quad \textit{for all } x \in B_\Phi,\ v \in B_{\Phi^*}.$$

**Proof.** Let $x \in C_\Phi$, $v \in C_{\Phi^*}$, then by (8)

$$\left|v(x)\right| \leqslant \Phi(x) + \Phi^*(v) \leqslant 2.$$

Let $x \in L_\Phi$, $v \in L_{\Phi^*}$. By Lemma 3.2, $x/\|x\|_\Phi \in C_\Phi$, $v/\|v\|_{\Phi^*} \in C_{\Phi^*}$ and therefore $|v(x)| \leqslant 2\|x\|_\Phi \|v\|_{\Phi^*}$. Clearly, the inequality extends to $x \in B_\Phi$, $v \in B_{\Phi^*}$. □

*3.3. The second conjugate*

If $E$ is a Banach space and $H \subset E$ is a closed subspace, then the dual of the quotient space $(E/H)$ can be identified with $H^\perp$. It follows that $\tilde{V}^* \cap V = (\mathrm{dom}(\Phi))^{\perp\perp}$, which is nothing else than the closure of $L_\Phi$ in $V$. Let us denote this space by $\bar{V}$.

As before, we can find the conjugate function to $\Phi^* : \tilde{V} \to \mathbb{R} \cup \{+\infty\}$ with respect to the pair $(\tilde{V}, \tilde{V}^*)$. Note that for $x$ in $\bar{V}$, we have

$$\sup_{v \in \tilde{V}} \left\{v(x) - \Phi^*(v)\right\} = \sup_{v \in V^*} \left\{v(x) - \Phi^*(v)\right\} = \Phi^{**}(x),$$

where $\Phi^{**}$ is the second conjugate to $\Phi : V \to \mathbb{R} \cup \{+\infty\}$. Since $\Phi$ is convex and lower semicontinuous, $\Phi^{**}(x) = \Phi(x)$ on $V$ [6]. It follows in particular that the restriction of $\Phi^{**}$ to $\bar{V}$ is a Young function.

It is clear from Hölder inequality that any $x \in L_\Phi$ defines a bounded linear functional on $B_{\Phi^*}$. Let $\|x\|^*_{\Phi^*}$ be its norm in $B^*_{\Phi^*}$, then by Lemma 3.2,

$$\|x\|^*_{\Phi^*} = \sup\left\{\left|v(x)\right|,\ \Phi^*(v) \leqslant 1\right\}.$$

Similarly, if $v \in L_{\Phi^*}$, then $v \in B^*_\Phi$ and we have

$$\|v\|^*_\Phi = \sup\left\{\left|v(x)\right|,\ \Phi(x) \leqslant 1\right\}.$$

**Lemma 3.6.** *For $x \in L_\Phi$, we have $\|x\|_\Phi \leqslant \|x\|^*_{\Phi^*} \leqslant 2\|x\|_\Phi$. Similarly, if $v \in L_{\Phi^*}$, then $\|v\|_{\Phi^*} \leqslant \|v\|^*_\Phi \leqslant 2\|v\|_{\Phi^*}$.*

**Proof.** Let $v \in L_{\Phi^*}$. By Hölder inequality, $\|v\|_\Phi^* \leqslant 2\|v\|_{\Phi^*}$. Let now $\|v\|_\Phi^* = 1$, then for $x \in C_\Phi$ we have

$$v(x) - \Phi(x) \leqslant 1.$$

On the other hand, for $x \in \mathrm{dom}(\Phi)$, such that $\Phi(x) > 1$, we get from Lemma 3.3

$$v(x) - \Phi(x) \leqslant v(x) - \|x\|_\Phi \leqslant 0.$$

It follows that $\Phi^*(v) \leqslant 1$ and $v \in C_{\Phi^*}$, hence $\|v\|_{\Phi^*} \leqslant 1$. Therefore, $\|v\|_{\Phi^*} \leqslant \|v\|_\Phi^*$ for all $v \in L_{\Phi^*}$. The proof for $x \in L_\Phi$ is the same, using the fact that $\Phi$ is the conjugate of $\Phi^*$. □

**Proposition 1.** *$B_{\Phi^*} \subseteq B_\Phi^*$ and $L_{\Phi^*} = \tilde{V} \cap B_\Phi^*$. Similarly, $B_\Phi \subseteq B_{\Phi^*}^*$ and $L_\Phi = \bar{V} \cap B_{\Phi^*}^*$.*

**Proof.** As we have seen, $L_{\Phi^*}$ is a vector subspace in $B_\Phi^*$ and the norms in $L_{\Phi^*}$ and $B_\Phi^*$ are equivalent, hence $B_{\Phi^*} \subseteq B_\Phi^*$. Let now $v \in \tilde{V} \cap B_\Phi^*$ be such that $\|v\|_\Phi^* = 1$. Then $\Phi^*(v) \leqslant 1$, exactly as in the proof of Lemma 3.6. It follows that for all $v \in \tilde{V} \cap B_\Phi^*$, $\Phi^*(v/\|v\|_\Phi^*) \leqslant 1 < \infty$ and $v \in L_{\Phi^*}$. Again, the proof for $L_\Phi$ and $B_\Phi$ is the same. □

Let $\Phi$ be a Young function such that 0 is an interior point in $\mathrm{dom}(\Phi)$. Then the function $\Phi$ is continuous in 0, therefore there is an open set $U$ containing 0 such that $U \subset C_\Phi$. It follows that $C_\Phi$ is a neighborhood of 0 in $V$, hence it is absorbing in $V$:

$$V = \bigcup_n nC_\Phi = L_\Phi \quad \text{(as sets)}. \tag{9}$$

Since $C_\Phi$ is a convex body (that is, 0 is a topological interior point), its Minkowski functional $\|\cdot\|_\Phi$ is continuous with respect to the original norm [14, p. 182]. It follows that we have the continuous inclusion $V \sqsubseteq B_\Phi$. Further, since $\mathrm{dom}(\Phi)$ has non-empty interior, $(\mathrm{dom}(\Phi))^\perp = \{0\}$ and $\tilde{V} = V^*$. Clearly also $\bar{V} = V$.

**Proposition 2.** *Let $0 \in \operatorname{int}\mathrm{dom}(\Phi)$. Then $V \sqsubseteq B_\Phi \subseteq B_{\Phi^*}^*$ and $L_{\Phi^*} = B_{\Phi^*} = B_\Phi^* \sqsubseteq V^*$.*

**Proof.** By (9), each $x \in V$ is in $L_\Phi$, and by continuity, $\|x\|_\Phi \leqslant K\|x\|$, for some fixed $K > 0$. Let $v \in B_\Phi^*$, then

$$\left|v(x)\right| \leqslant \|v\|_\Phi^* \|x\|_\Phi \leqslant K\|v\|_\Phi^* \|x\| \quad \text{for } x \in V,$$

hence $v \in V^* = \tilde{V}$ and $\|v\|^* \leqslant K\|v\|_\Phi^*$. The statement now follows from Proposition 1. □

## 4. The spaces $B_\varphi$ and $B_{\varphi,0}$

Let $\mathcal{M}_s$ be the real Banach subspace of self-adjoint elements in $\mathcal{M}$, then the dual $\mathcal{M}_s^*$ is the subspace of Hermitian (not necessarily normal) functionals in $\mathcal{M}^*$. We define the functional $F_\varphi : \mathcal{M}_s^* \to \mathbb{R} \cup \{\infty\}$ by

$$F_\varphi(\omega) = \begin{cases} S(\omega, \varphi) & \text{if } \omega \in \mathfrak{S}_*, \\ \infty & \text{otherwise.} \end{cases}$$

Then $F_\varphi$ is convex and lower semicontinuous, with $\operatorname{dom}(F_\varphi)=\mathcal{S}_\varphi$. It follows from (1) that $F_\varphi$ is strictly convex. Its conjugate $F_\varphi^*$ is

$$F_\varphi^*(h)=\sup_{\omega\in\mathfrak{S}_*}\bigl\{\omega(h)-F_\varphi(\omega)\bigr\}=c_\varphi(h),\quad h\in\mathcal{M}_s.$$

Hence $c_\varphi$ is convex and lower semicontinuous, in fact, since finite valued, it is continuous on $\mathcal{M}_s$. We have $c_\varphi^*=F_\varphi^{**}=F_\varphi$ on $\mathcal{M}_s^*$. Note also that

$$c_\varphi(h+\lambda)=c_\varphi(h)+\lambda,\quad \forall\lambda\in\mathbb{R}. \tag{10}$$

We define another convex and lower semicontinuous functional on $\mathcal{M}_s^*$, namely,

$$\bar F_\varphi(\omega)=\begin{cases} S(\omega,\varphi)-\omega(1) & \text{if } \omega\in\mathcal{M}_*^+,\\ \infty & \text{otherwise.}\end{cases}$$

Then the conjugate functional is

$$\begin{aligned}\bar F_\varphi^*(h)&=\sup_{\omega\in\mathcal{M}_*^+}\bigl\{\omega(h)-S(\omega,\varphi)+\omega(1)\bigr\}=\sup_{\omega\in\mathfrak{S}_*,\lambda\in\mathbb{R}^+}\bigl\{\lambda\omega(h)-S(\lambda\omega,\varphi)+\lambda\bigr\}\\ &=\sup_{\omega\in\mathfrak{S}_*,\lambda\in\mathbb{R}^+}\bigl\{\lambda\bigl(\omega(h)-S(\omega,\varphi)\bigr)-\lambda\log\lambda+\lambda\bigr\}\\ &=\sup_{\lambda\in\mathbb{R}^+}\bigl\{\lambda\bigl(c_\varphi(h)+1\bigr)-\lambda\log\lambda\bigr\}=e^{c_\varphi(h)}=\varphi^h(1).\end{aligned}$$

Again, $h\mapsto\varphi^h(1)$ is convex and continuous and $\bar F_\varphi^{**}=\bar F_\varphi$.

Next, we define a Young function on $\mathcal{M}_s$. Let $\Phi_\varphi:\mathcal{M}_s\to\mathbb{R}^+$ be defined by

$$\Phi_\varphi(h)=\frac{\varphi^h(1)+\varphi^{-h}(1)}{2}-1.$$

**Lemma 4.1.** *$\Phi_\varphi$ is a Young function.*

**Proof.** The property (i) from the definition of a Young function follows from the properties of $h\mapsto\varphi^h(1)$. Since $\varphi^h(1)=e^{c_\varphi(h)}\geqslant e^{\omega(h)-S(\omega,\varphi)}$ for all normal states $\omega$, we have

$$\Phi_\varphi(h)\geqslant\cosh\bigl(\omega(h)\bigr)e^{-S(\omega,\varphi)}-1. \tag{11}$$

In particular,

$$\Phi_\varphi(h)\geqslant\cosh\bigl(\varphi(h)\bigr)-1\geqslant 0\quad\text{for all } h. \tag{12}$$

Since obviously $\Phi_\varphi(0)=0$, (ii) follows. Let now $h$ be such that $\omega(h)=0$ for all $\omega\in\mathcal{S}_\varphi$, then by definition, $c_\varphi(h)=0$ and $\varphi=\varphi^h$, hence $h=0$. Therefore if $h\neq 0$, then there is a state $\omega\in\mathcal{S}_\varphi$ such that $\omega(h)\neq 0$ and then $\lim_{t\to\infty}\cosh(t\omega(h))=\infty$, this implies (iv). Property (iii) is obviously satisfied. □

Let $C_\varphi := C_{\Phi_\varphi}$, $B_\varphi := B_{\Phi_\varphi}$ and $\|\cdot\|_\varphi := \|\cdot\|_{\Phi_\varphi}$. Since $\operatorname{dom}\Phi_\varphi = \mathcal{M}_s$, we have by Proposition 2 that $\mathcal{M}_s \sqsubseteq B_\varphi$. If $\Phi_\varphi^*$ is the conjugate of $\Phi_\varphi$, then $B_\varphi^* = B_{\Phi_\varphi^*} \sqsubseteq \mathcal{M}_s^*$.

Let now $h \in \mathcal{M}_s$, such that $\|h\|_\varphi = t > 0$, that is,

$$\Phi_\varphi\left(\frac{h}{t}\right) = 1.$$

If $\omega$ is a state, then by (11),

$$\cosh\left(\frac{\omega(h)}{t}\right) \leqslant 2e^{S(\omega,\varphi)}. \tag{13}$$

If $\omega \in \mathcal{S}_\varphi$, then $|\omega(h)| \leqslant ct$, where $c > 0$ is some constant depending on $S(\omega, \varphi)$. It follows that each $\omega \in \mathcal{S}_\varphi$ extends to a continuous linear functional on $B_\varphi$. Moreover, for $C > 0$, $\mathcal{S}_{\varphi,C}$ is an equicontinuous subset in $B_\varphi^*$.

Let $\mathcal{M}_{s,0} \subset \mathcal{M}_s$ be the subspace of elements satisfying $\varphi(h) = 0$. Then by putting $\omega = \varphi$ in (6), we get

$$c_\varphi(h) = S\big(\varphi, \big[\varphi^h\big]\big) \geqslant 0.$$

Let us define

$$\Phi_{\varphi,0}(h) = \frac{c_\varphi(h) + c_\varphi(-h)}{2}, \quad h \in \mathcal{M}_{\varphi,0}.$$

Then it is easy to check that $\Phi_{\varphi,0}$ is a Young function on $\mathcal{M}_{\varphi,0}$. We have

**Lemma 4.2.** *Let $h \in \mathcal{M}_{s,0}$. Then*

$$\Phi_{\varphi,0}(h) \leqslant \Phi_\varphi(h) \leqslant e^{2\Phi_{\varphi,0}} - 1.$$

**Proof.** The first inequality follows from $a \leqslant e^a - 1$ for $a \geqslant 0$, the second follows from $x + y \leqslant 2xy$ for $x, y \geqslant 1$. □

Let us construct the Banach space $B_{\Phi_{\varphi,0}} =: B_{\varphi,0}$ and let $\|\cdot\|_{\varphi,0} := \|\cdot\|_{\Phi_{\varphi,0}}$.

**Proposition 3.** *The norms $\|\cdot\|_{\varphi,0}$ and $\|\cdot\|_\varphi$ are equivalent on $\mathcal{M}_{s,0}$.*

**Proof.** Let us denote $C_{\varphi,0} := C_{\Phi_{\varphi,0}}$. We show that

$$\frac{1}{2}\log 2\, C_{\varphi,0} \subseteq C_\varphi \cap \mathcal{M}_{s,0} \subseteq C_{\varphi,0}. \tag{14}$$

Let $h \in C_{\varphi,0}$ and $t = \frac{1}{2}\log 2$. Then by convexity, $\Phi_{\varphi,0}(th) \leqslant t = \frac{1}{2}\log 2$ and hence

$$\Phi_\varphi(th) \leqslant e^{2\Phi_{\varphi,0}(th)} - 1 \leqslant 1,$$

which implies $tC_{\varphi,0} \subseteq C_\varphi \cap \mathcal{M}_{s,o}$. The other inclusion follows from the first inequality in Lemma 4.2. It follows from (14) that for $h \in \mathcal{M}_{s,0}$,

$$\|h\|_{\varphi,0} \leqslant \|h\|_\varphi \leqslant \frac{2}{\log 2}\|h\|_{\varphi,0}. \qquad \square$$

Note that since $\varphi \in \mathcal{S}_\varphi$, $\varphi$ extends to a bounded linear functional on $B_\varphi$. Clearly, the completion of $\mathcal{M}_{s,0}$ under the norm $\|\cdot\|_\varphi$ is the Banach subspace $\{h \in B_\varphi,\ \varphi(h) = 0\}$. It follows from the above proposition that $B_{\varphi,0}$ can be identified with the subspace of centered elements in $B_\varphi$.

## 5. Extension of $c_\varphi$

Since $\mathcal{S}_\varphi \subset B_\varphi^* \sqsubseteq \mathcal{M}_s^*$, the restriction of $F_\varphi$ is a strictly convex lower semicontinuous functional on $B_\varphi^*$, with effective domain $S_\varphi$. Its conjugate $F_\varphi^*$ is a lower semicontinuous extension of $c_\varphi$ to $B_\varphi$, moreover, $F_\varphi^{**} = F_\varphi$. We show that this extension has again values in $\mathbb{R}$ and is continuous.

**Lemma 5.1.** *Let the sequence $\{h_n\}_n \subset \mathcal{M}_s$ be Cauchy in the norm $\|\cdot\|_\varphi$. Then the sequences $\{c_\varphi(h_n)\}_n$ and $\{S([\varphi^{h_n}], \varphi)\}_n$ are bounded.*

**Proof.** By (5), we have for all $n$

$$\varphi(h_n) \leqslant c_\varphi(h_n).$$

Since $\varphi(h_n)$ converges, $c_\varphi(h_n)$ is bounded from below. Let $n_0$ be such that $\|h_n - h_{n_0}\|_\varphi < 1$ for all $n \geqslant n_0$, then

$$\omega(h_n) - S(\omega, \varphi) \leqslant \omega(h_{n_0}) + c_\varphi(h_n - h_{n_0}) \leqslant \|h_{n_0}\| + \log 2$$

for all such $n$ and $\omega \in S_\varphi$. It follows that $\{c_\varphi(h_n)\}_n$ is bounded.

If $\{h_n\}_n$ is Cauchy, then the sequence $\{th_n\}_n$ is also Cauchy for all $t \in \mathbb{R}$ and there are constants $A_t$, $B_t$, such that

$$A_t \leqslant c_\varphi(th_n) \leqslant B_t, \quad \forall n.$$

On the other hand, we have

$$\frac{d}{dt}c_\varphi(th_n)\bigg|_{t=1} = \big[\varphi^{h_n}\big](h_n).$$

By convexity,

$$c_\varphi(th_n) \geqslant c_\varphi(h_n) + (t-1)\frac{d}{dt}c_\varphi(th_n)\bigg|_{t=1} \geqslant A_1 + (t-1)\big[\varphi^{h_n}\big](h_n).$$

For arbitrary fixed $t > 1$, we get

$$\big[\varphi^{h_n}\big](h_n) \leqslant \frac{B_t - A_1}{t-1}, \quad \forall n.$$

Boundedness of $S([\varphi^{h_n}],\varphi)$ now follows from

$$0 \leqslant S\big(\big[\varphi^{h_n}\big],\varphi\big) = \big[\varphi^{h_n}\big](h_n) - c_\varphi(h_n). \qquad \square$$

**Theorem 4.** *Let $\{h_n\}_n$ be a sequence in $\mathcal{M}_s$, converging to some $h$ in $B_\varphi$. Then*

$$\lim_n c_\varphi(h_n) = \sup_{\omega\in\mathcal{S}_\varphi} \big\{\omega(h) - S(\omega,\varphi)\big\} \tag{15}$$

*and there is a unique state $\psi \in \mathcal{S}_\varphi$ such that the supremum is attained. The state $\psi$ is faithful. Moreover,* $\lim_n S([\varphi^{h_n}],\varphi) = S(\psi,\varphi)$, $\lim_n[\varphi^{h_n}(h_n)] = \psi(h)$ *and* $\lim_n S(\psi,[\varphi^{h_n}]) = 0$. *In particular, $[\varphi^{h_n}]$ converges to $\psi$ in norm.*

The state $\psi$ will be denoted by $[\varphi^h]$ and the limit $\lim_n c_\varphi(h_n) =: c_\varphi(h)$.

**Proof.** This proof is similar to the proof of [16, Theorem 12.3].

By Lemma 5.1, there is some $C > 0$ such that $[\varphi^{h_n}] \in \mathcal{S}_{\varphi,C}$ for all $n$. The set $\mathcal{S}_{\varphi,C}$ is weakly relatively compact in $\mathfrak{S}_*$ and hence there is subsequence $[\varphi^{h_{n_k}}]$ converging weakly to a state $\psi \in \mathcal{S}_{\varphi,C}$. We will show that $[\varphi^{h_{n_k}}](h_{n_k})$ converges to $\psi(h)$.

Since $\mathcal{S}_{\varphi,C}$ is an equicontinuous subset in $B_\varphi^*$, $\omega(h_n)$ converges to $\omega(h)$ uniformly for all $\omega \in \mathcal{S}_{\varphi,C}$. This implies

$$\big|\big[\varphi^{h_{n_k}}\big](h_{n_k}) - \big[\varphi^{h_{n_k}}\big](h)\big| < \varepsilon$$

for sufficiently large $n_k$. We further have

$$\begin{aligned}\big|\big[\varphi^{h_{n_k}}\big](h) - \psi(h)\big| &\leqslant \big|\big[\varphi^{h_{n_k}}\big](h) - \big[\varphi^{h_{n_k}}\big](h_m)\big| + \big|\big[\varphi^{h_{n_k}}\big](h_m) - \psi(h_m)\big| \\ &\quad + \big|\psi(h_m) - \psi(h)\big| < \varepsilon\end{aligned}$$

for sufficiently large $m$ and $n_k$. Putting both inequalities together, we get $[\varphi^{h_{n_k}}](h_{n_k}) \to \psi(h)$.

Let $\omega \in \mathcal{S}_\varphi$. By definition,

$$\big[\varphi^{h_{n_k}}\big](h_{n_k}) - S\big(\big[\varphi^{h_{n_k}}\big],\varphi\big) = c_\varphi(h_{n_k}) \geqslant \omega(h_{n_k}) - S(\omega,\varphi).$$

By weak lower semicontinuity of the relative entropy, we get

$$\psi(h) - S(\psi,\varphi) \geqslant \limsup c_\varphi(h_{n_k}) \geqslant \omega(h) - S(\omega,\varphi) \tag{16}$$

and thus $\psi$ is a maximizer of (15). On the other hand,

$$\psi(h_{n_k}) - S(\psi,\varphi) \leqslant \big[\varphi^{h_{n_k}}\big](h_{n_k}) - S\big(\big[\varphi^{h_{n_k}}\big],\varphi\big) = c_\varphi(h_{n_k}).$$

From this and (16), it follows that $\psi(h) - S(\psi,\varphi) = \lim c_\varphi(h_{n_k})$. It also follows that

$$\limsup S\big(\big[\varphi^{h_{n_k}}\big],\varphi\big) \leqslant S(\psi,\varphi)$$

and this, together with lower semicontinuity implies that $S([\varphi^{h_{n_k}}],\varphi)$ converges to $S(\psi,\varphi)$.

To show that such $\psi$ is unique, suppose that $\psi'$ is another maximizer, then for $\psi_\lambda := \lambda\psi + (1-\lambda)\psi'$, $0 \leqslant \lambda \leqslant 1$, we have

$$\begin{aligned}\psi(h) - S(\psi,\varphi) &\geqslant \psi_\lambda(h) - S(\psi_\lambda,\varphi)\\ &\geqslant \psi_\lambda(h) - \lambda S(\psi,\varphi) - (1-\lambda)S(\psi',\varphi) = \psi(h) - S(\psi,\varphi)\end{aligned}$$

hence $\psi_\lambda$ is a maximizer as well and, moreover,

$$S(\psi_\lambda,\varphi) = \lambda S(\psi,\varphi) + (1-\lambda)S(\psi',\varphi).$$

By strict convexity, this implies that $\psi = \psi'$. It also follows that the whole sequence $[\varphi^{h_n}]$ converges weakly to $\psi$.

Using (6), we have

$$S(\varphi,\psi) \leqslant \liminf_n S\big(\varphi,\big[\varphi^{h_n}\big]\big) = \lim_n c_\varphi(h_n) - \varphi(h) < \infty.$$

This implies that $\operatorname{supp}\varphi \leqslant \operatorname{supp}\psi$ and $\psi$ is faithful. Finally, by taking the limit in the equality

$$\psi(h_n) - S(\psi,\varphi) = c_\varphi(h_n) - S\big(\psi,\big[\varphi^{h_n}\big]\big)$$

we get $\lim_n S(\psi,[\varphi^{h_n}]) \to 0$. □

**Corollary 5.1.** *Let $h_n$ be a sequence in $B_\varphi$, then $h_n \to 0$ if and only if $c_\varphi(th_n) \to 0$ for all $t \in \mathbb{R}$.*

**Proof.** Let $h_n$ be such that $c_\varphi(th_n) = \log\varphi^{th_n}(1)$ converges to 0, then $\varphi^{th_n}(1)$ converges to 1, for all $t \in \mathbb{R}$. Therefore, for each $\varepsilon > 0$, $\Phi_\varphi(h_n/\varepsilon) < 1$ for large enough $n$, that is, $\|h_n\|_\varphi \to 0$. The converse follows from Theorem 4. □

In particular, if $h_n \in \mathcal{M}_s$ is a sequence converging strongly to $h$, then $h_n$ converges to $h$ in $\|\cdot\|_\varphi$, see [16].

## 6. The dual spaces

The dual space $\mathcal{M}^*_{s,0}$ is obtained as the quotient space $\mathcal{M}^*_s/\{\varphi\}$. Each equivalence class in $\mathcal{M}^*_{s,0}$ can be identified with its unique element $v$ satisfying $v(1) = 0$. By Proposition 2, we have $B^*_{\varphi,0} = B_{\Phi^*_{\varphi,0}} \sqsubseteq \mathcal{M}^*_{s,0}$. By Proposition 3, $B^*_{\varphi,0}$ is the same as $B^*_\varphi/\{\varphi\}$.

**Lemma 6.1.** *Let $\bar{c}_\varphi$ be the restriction of $c_\varphi$ to $B_{\varphi,0}$. Then the conjugate functional is $\bar{c}^*_\varphi(v) = F_\varphi(v+\varphi)$.*

**Proof.** Let $v \in B^*_\varphi$, $v(1) = 0$. Then by (10),

$$\begin{aligned}F_\varphi(v+\varphi) &= \sup_{h\in B_\varphi}\big\{v(h) + \varphi(h) - c_\varphi(h)\big\}\\ &= \sup_{h\in B_\varphi}\big\{v\big(h-\varphi(h)\big) - \bar{c}_\varphi\big(h-\varphi(h)\big)\big\} = \bar{c}^*_\varphi(v). \qquad \Box\end{aligned}$$

Let $V$ be a Banach space and $V^*$ its dual. For any subset $D \subset V$, let $D^\circ$ be the polar of $D$ in $V^*$, that is, $D^\circ = \{v \in V^*,\ v(h) \leqslant 1,\ \forall h \in D\}$. We will need the following lemma.

**Lemma 6.2.** *Let $F : V \to \mathbb{R}^+$ be a convex functional such that $F(0) = 0$ and let $F^*$ be its conjugate. Let $D = \{x \in V,\ F(x) \leqslant 1\}$ and $D^* = \{v \in V^*,\ F^*(v) \leqslant 1\}$. Then*

$$\frac{1}{2} D^* \subseteq D^\circ \subseteq D^*.$$

**Proof.** If $v \in D^*$, then $v(x) \leqslant F(x) + F^*(v) \leqslant 2$ for all $x \in D$ and therefore $\frac{1}{2}v \in D^\circ$. Let now $v \in D^\circ$, then

$$v(x) - 1 \leqslant 0 \leqslant F(x) \quad \text{for } x \in D.$$

If $F(x) > 1$, then by continuity there is some $t \in (0,1)$ such that $F(tx) = 1$. Since $tx \in D$, $v(tx) \leqslant 1$, moreover, by convexity, $1 = F(tx) \leqslant tF(x)$. Consequently,

$$v(x) - 1 \leqslant \frac{1}{t} - 1 \leqslant F(x).$$

It follows that $F^*(v) \leqslant 1$ and $v \in D^*$. □

Let us denote $K_{\varphi,0} := \{h \in B_{\varphi,0},\ \Phi_{\varphi,0}(h) \leqslant 1\}$. Then $K_{\varphi,0}$ is the closed unit ball in $B_{\varphi,0}$. Its polar $K^\circ_{\varphi,0}$ is the closed unit ball in $B^*_{\varphi,0}$.

**Proposition 5.** *Let $v$ be an element in $K^\circ_{\varphi,0}$. Then there are states $\omega_1$, $\omega_2$, satisfying $S(\omega_1, \varphi) + S(\omega_2, \varphi) \leqslant 1$, such that $v = \omega_1 - \omega_2$.*

**Proof.** Since $\bar{c}_\varphi$ is continuous on $B_{\varphi,0}$, the set $D := \{h \in B_{\varphi,0},\ \bar{c}_\varphi(h) \leqslant 1\}$ is closed. Let us endow the dual pair $B_{\varphi,0}$ and $B^*_{\varphi,0}$ with the $\sigma(B_{\varphi,0}, B^*_{\varphi,0})$ and $\sigma(B^*_{\varphi,0}, B_{\varphi,0})$ topology, respectively. As $D$ is convex, it is closed also in this weaker topology. The set $D \cap -D$ is absolutely convex and closed, moreover,

$$D \cap -D \subseteq K_{\varphi,0} \subseteq 2(D \cap -D), \tag{17}$$

as can be easily checked. Then

$$\frac{1}{2}(D \cap -D)^\circ \subseteq K^\circ_{\varphi,0} \subseteq (D \cap -D)^\circ.$$

By [14], $(D \cap -D)^\circ$ is the closed convex cover of $D^\circ \cup -D^\circ$, which is the same as the closed absolutely convex cover of $D^\circ$. Moreover, since $D^\circ$ is the polar of a neighborhood of 0, it is compact [14]. Therefore its absolutely convex cover is also compact, hence closed. It follows that $(D \cap -D)^\circ$ is the absolutely convex cover of $D^\circ$.

By Lemmas 6.1 and 6.2,

$$\frac{1}{2}(\mathcal{S}_{\varphi,1} - \varphi) \subseteq D^\circ \subseteq \mathcal{S}_{\varphi,1} - \varphi$$

and this implies

$$\frac{1}{4}\operatorname{abs\,conv}(\mathcal{S}_{\varphi,1}-\varphi) \subseteq K^{\circ}_{\varphi,0} \subseteq \operatorname{abs\,conv}(\mathcal{S}_{\varphi,1}-\varphi). \tag{18}$$

Let now $v \in \operatorname{abs\,conv}(\mathcal{S}_{\varphi,1}-\varphi)$, then there are elements $\varphi_1,\ldots,\varphi_n \in \mathcal{S}_{\varphi,1}$, and real numbers $\lambda_1,\ldots,\lambda_n$, $\sum_n |\lambda_n| = 1$, such that $v = \sum_n \lambda_n(\varphi_n - \varphi)$. Let $m \leqslant n$ be such that $\lambda_i > 0$ for $i \leqslant m$ and $\lambda_i < 0$ for $i > m$. Then $v = \omega_1 - \omega_2$, with

$$\omega_1 = \sum_{i=1}^{m} \lambda_i\varphi_i + (1-\lambda)\varphi, \qquad \omega_2 = \sum_{i=m+1}^{n} |\lambda_i|\varphi_i + \lambda\varphi,$$

where $\lambda = \sum_{i=1}^{m}\lambda_i$. Moreover, $S(\omega_1,\varphi) \leqslant \sum_i^m \lambda_i S(\varphi_i,\varphi) \leqslant \lambda$, and similarly, $S(\omega_2,\varphi) \leqslant 1-\lambda$. □

**Theorem 6.**

(i) $B^*_\varphi = \mathcal{P}_\varphi - \mathcal{P}_\varphi$ *and* $B^*_\varphi \cap \mathcal{M}^+_* = \mathcal{P}_\varphi$.
(ii) $B^*_{\varphi,0} = \bigcup_n n(\mathcal{S}_{\varphi,1} - \mathcal{S}_{\varphi,1})$.

**Proof.** (i) Let $\omega \in B^*_\varphi$ and let $v = \omega - \omega(1)\varphi$. Then $v$ can be seen as an element in $B^*_{\varphi,0}$. Let $\|v\|^*_{\varphi,0} = t$, then by Proposition 5, there are $\omega_1,\omega_2 \in \mathcal{S}_{\varphi,1}$, such that $v/t = \omega_1 - \omega_2$, that is, $\omega = t\omega_1 + \omega(1)\varphi - t\omega_2$. Since $\omega_1,\omega_2,\varphi \in \mathcal{P}_\varphi$ and $\mathcal{P}_\varphi$ is a convex cone, it follows that $B^*_\varphi \subseteq \mathcal{P}_\varphi - \mathcal{P}_\varphi$. On the other hand, we have already shown that if $\omega \in \mathcal{S}_\varphi$, then $\omega \in B^*_\varphi$ and hence $\mathcal{P}_\varphi - \mathcal{P}_\varphi \subseteq B^*_\varphi$. Let $\omega \in B^*_\varphi \cap \mathcal{M}^+_*$, then we get $\omega + t\omega_2 = t\omega_1 + \omega(1)\varphi$. It follows that $\omega + t\omega_2 \in \mathcal{P}_\varphi$, and identity (1) implies that $\omega$ must be in $\mathcal{P}_\varphi$.

(ii) By Proposition 5,

$$K^{\circ}_{\varphi,0} \subseteq (\mathcal{S}_{\varphi,1} - \mathcal{S}_{\varphi,1}) \subseteq 4K^{\circ}_{\varphi,0}.$$

The equality now follows from the fact that the closed unit ball is absorbing in $B^*_{\varphi,0}$. □

In the rest of this section, we find an equivalent norm on $B^*_{\varphi,0}$. We define a function $f : \mathfrak{S}_* \times \mathfrak{S}_* \to \mathbb{R}^+$ by

$$f(\omega_1,\omega_2) = S(\omega_1,\varphi) + S(\omega_2,\varphi).$$

Clearly, $f$ is weakly lower semicontinuous and strictly convex. Further, let $v \in \mathfrak{S}_* - \mathfrak{S}_*$ and let $L_v = \{(\omega_1,\omega_2) \in \mathfrak{S}_* \times \mathfrak{S}_*,\ \omega_1 - \omega_2 = v\}$. Then $L_v$ is a weakly closed subset in $\mathcal{M}_* \times \mathcal{M}_*$.

**Lemma 6.3.** *Let $v \in \mathcal{S}_\varphi - \mathcal{S}_\varphi$. Then the function $f$ attains its minimum over $L_v$ at a unique point $(v_+, v_-) \in L_v$.*

**Proof.** By assumptions, $v = \omega_1 - \omega_2$ for some $\omega_1,\omega_2 \in \mathcal{S}_\varphi$. Let $C > 0$ be such that $\omega_1,\omega_2 \in \mathcal{S}_{\varphi,C}$, then the infimum is taken over the set $L_v \cap \mathcal{S}_{\varphi,C} \times \mathcal{S}_{\varphi,C}$. Since $L_v$ is weakly closed and $\mathcal{S}_{\varphi,C}$ is weakly compact, the intersection is weakly compact and $f$ attains its minimum on it. Uniqueness follows by strict convexity of $f$. □

Let us now define the functional $\Psi_{\varphi,0} : \mathcal{M}^*_{s,0} \to \mathbb{R}^+$ by

$$\Psi_{\varphi,0}(v) = \begin{cases} f(v_+, v_-) & \text{if } v \in \mathcal{S}_\varphi - \mathcal{S}_\varphi, \\ \infty & \text{otherwise.} \end{cases}$$

**Lemma 6.4.** *$\Psi_{\varphi,0}$ is a Young function.*

**Proof.** It is easy to check that $\Psi_{\varphi,0}$ is convex, positive, $\Psi_{\varphi,0}(v) = \Psi_{\varphi,0}(-v)$ and that $\Psi_{\varphi,0}(v) = 0$ if and only if $v = 0$. We will show that $\Psi_{\varphi,0}$ is lower semicontinuous.

To do this, we have to prove that for any $C > 0$, the set of all $v$ satisfying $\Psi_{\varphi,0}(v) \leqslant C$ is closed. Let $v_n$ be a sequence of elements in this set, converging to some $v$. Let $v_n = v_{n+} - v_{n-}$ be the corresponding decompositions, then $v_{n+}, v_{n-} \in \mathcal{S}_{\varphi,C}$ for all $n$, hence there are elements $v'_+$ and $v'_-$ in $\mathcal{S}_{\varphi,C}$ and a subsequence $v_{n_k} = v_{n_k+} - v_{n_k-}$ such that $v_{n_k+} \to v'_+$ and $v_{n_k-} \to v'_-$ weakly. It follows that $v = v'_+ - v'_-$ and $\Psi_{\varphi,0}(v) \leqslant S(v'_+, \varphi) + S(v'_-, \varphi) \leqslant \liminf S(v_{n_k+}, \varphi) + S(v_{n_k-}, \varphi) \leqslant C$.

Suppose that $v \neq 0$, then $\Psi_{\varphi,0}(v) > 0$. If $t > 1$, then by convexity, $t\Psi_{\varphi,0}(v) \leqslant \Psi_{\varphi,0}(tv)$, hence $\lim_{t\to\infty} \Psi_{\varphi,0}(tv) = \infty$. □

Let us find the corresponding Banach space. Note that

$$C_{\Psi_{\varphi,0}} = \bigl\{\omega_1 - \omega_2\colon\ \omega_1, \omega_2 \in \mathfrak{S}_*,\ S(\omega_1, \varphi) + S(\omega_2, \varphi) \leqslant 1\bigr\}.$$

By Proposition 5, this implies that $K^\circ_{\varphi,0} \subseteq C_{\Psi_{\varphi,0}} \subseteq \mathcal{S}_{\varphi,1} - \mathcal{S}_{\varphi,1}$ and by Theorem 6(ii), $B^*_{\varphi,0} \subseteq L_{\Psi_{\varphi,0}} \subseteq B^*_{\varphi,0}$.

**Proposition 7.** *$\|\cdot\|_{\Psi_{\varphi,0}}$ defines an equivalent norm in $B^*_{\varphi,0}$.*

**Proof.** Let $\Psi^*_{\varphi,0} : \mathcal{M}_s \to \mathbb{R}$ be the conjugate functional, then

$$\begin{aligned} \Psi^*_{\varphi,0}(h) &= \sup_{v \in \mathcal{M}^*_{s,0}} v(h) - \Psi_{\varphi,0}(v) \\ &= \sup_{v \in \mathcal{S}_\varphi - \mathcal{S}_\varphi} \sup_{(\omega_1,\omega_2) \in L_v} \omega_1(h) - \omega_2(h) - f(\omega_1, \omega_2) \\ &= \sup_{\omega_1,\omega_2 \in \mathcal{S}_\varphi} \omega_1(h) - S(\omega_1, \varphi) + \omega_2(-h) - S(\omega_2, \varphi) = 2\Phi_{\varphi,0}(h). \end{aligned}$$

It follows that $\Psi_{\varphi,0}(v) = \Psi^{**}_{\varphi,0}(v) = 2\Phi^*_{\varphi,0}(\frac{1}{2}v)$. Since the norms $\|\cdot\|^*_{\varphi,0}$ and $\|\cdot\|_{\Phi^*_{\varphi,0}}$ are equivalent, this finishes the proof. □

## 7. The chain rule

**Proposition 8.** *Let $h \in B_\varphi$, $k \in \mathcal{M}_s$. Then $[\varphi^{h+k}] = [[\varphi^h]^k]$, $c_\varphi(h+k) = c_{[\varphi^h]}(k) + c_\varphi(h)$ and for all normal states $\omega$ the equality*

$$\omega(k) - S\bigl(\omega, [\varphi^h]\bigr) = c_\varphi(h+k) - c_\varphi(h) - S\bigl(\omega, [\varphi^{h+k}]\bigr) \tag{19}$$

*holds.*

**Proof.** Let $h_n \in \mathcal{M}_s$ be such that $h_n \to h$ in $B_\varphi$. By the chain rule (7), we have $[\varphi^{h_n+k}] = [[\varphi^{h_n}]^k]$ and $c_\varphi(h_n + k) = c_{[\varphi^{h_n}]}(k) + c_\varphi(h_n)$. By Theorem 4, $c_\varphi(h_n) \to c_\varphi(h)$, $c_\varphi(h_n + k) \to c_\varphi(h + k)$ and $[\varphi^{h_n}] \to [\varphi^h]$, $[\varphi^{h_n+k}] \to [\varphi^{h+k}]$ strongly. Now we can proceed exactly as in the proof of [16, Theorem 12.10] to obtain (19). By putting $\omega = [\varphi^{h+k}]$ in this equality, we get

$$\big[\varphi^{h+k}\big](k) + S\big(\big[\varphi^{h+k}\big], \big[\varphi^h\big]\big) = c_\varphi(h+k) - c_\varphi(h) \geqslant \omega(k) - S\big(\omega, \big[\varphi^h\big]\big)$$

for all $\omega$, which implies the statement of the proposition. □

**Theorem 9.** *Let $h \in B_\varphi$. Then $B_{[\varphi^h]} = B_\varphi$ and $\mathcal{S}_{[\varphi^h]} = \mathcal{S}_\varphi$.*

**Proof.** Let $k \in \mathcal{M}_s$ and let $\varepsilon > 0$. By Proposition 8,

$$c_{[\varphi^h]}(k) = c_\varphi(h+k) - c_\varphi(h).$$

Since $c_\varphi$ is continuous on $B_\varphi$, there is a $\delta > 0$ such that

$$\big|c_\varphi(h+k) - c_\varphi(h)\big| < \log 2$$

if $\|k\|_\varphi < \delta$. It follows that $\|k\|_{[\varphi^h]} < \varepsilon$ whenever $\|k\|_\varphi < \delta\varepsilon$ and this implies $B_\varphi \sqsubseteq B_{[\varphi^h]}$. In particular, $h \in B_{[\varphi^h]}$.

Let $h_n$ be a sequence converging to $h$ in $B_\varphi$, then by (6)

$$\omega(h_n) - S(\omega, \varphi) = c_\varphi(h_n) - S\big(\omega, \big[\varphi^{h_n}\big]\big).$$

By Theorem 4, and lower semicontinuity,

$$\omega(h) - S(\omega, \varphi) \leqslant c_\varphi(h) - S\big(\omega, \big[\varphi^h\big]\big).$$

This implies $\mathcal{S}_\varphi \subseteq \mathcal{S}_{[\varphi^h]}$.

Further, $h_n$ converges to $h$ in $B_{[\varphi^h]}$ and by Theorem 4 and Proposition 8,

$$\big[\big[\varphi^h\big]^{-h}\big] = \lim_n \big[\big[\varphi^h\big]^{-h_n}\big] = \lim_n \big[\varphi^{h-h_n}\big] = \varphi.$$

By the first part of the proof, $B_{[\varphi^h]} = B_\varphi$ and $\mathcal{S}_\varphi = \mathcal{S}_{[\varphi^h]}$. □

**Theorem 10.** *Let $h, k \in B_\varphi$. Then the chain rule $c_\varphi(h+k) = c_{[\varphi^h]}(k) + c_\varphi(h)$, $[[\varphi^h]^k] = [\varphi^{h+k}]$ holds.*

**Proof.** Let $k_n \in \mathcal{M}_s$ be a sequence converging to $k$ in $B_\varphi = B_{[\varphi^h]}$. Then

$$\big[\big[\varphi^h\big]^k\big] = \lim_n \big[\big[\varphi^h\big]^{k_n}\big] = \lim_n \big[\varphi^{h+k_n}\big] = \big[\varphi^{h+k}\big]$$

and by Proposition 8,

$$c_\varphi(h+k) = \lim_n c_{[\varphi^h]}(k_n) + c_\varphi(h) = c_{[\varphi^h]}(k) + c_\varphi(h). \qquad \square$$

**Corollary 7.1.** *Let $h \in B_\varphi$ and let $\omega$ be a normal state. Then the equality*

$$\omega(h) - S(\omega, \varphi) = c_\varphi(h) - S\big(\omega, \big[\varphi^h\big]\big)$$

*holds.*

**Proof.** By (6) and lower semicontinuity, we have

$$\omega(h) - S(\omega, \varphi) \leqslant c_\varphi(h) - S\big(\omega, \big[\varphi^h\big]\big).$$

Since, by the chain rule, $\varphi = [[\varphi^h]^{-h}]$ and $c_{[\varphi^h]}(-h) = -c_\varphi(h)$, we also have

$$\omega(-h) - S\big(\omega, \big[\varphi^h\big]\big) \leqslant c_{[\varphi^h]}(-h) - S(\omega, \varphi) = -c_\varphi(h) - S(\omega, \varphi)$$

which implies the opposite inequality. □

**Corollary 7.2.** *Let $[\varphi^h] = [\varphi^k]$ for some $h, k \in B_\varphi$. Then $h - k = \varphi(h - k)$.*

**Proof.** Let us suppose that $h \in B_\varphi$ is such that $[\varphi^h] = \varphi$. Then $[\varphi^{nh}] = \varphi$ for all $n \in \mathbb{N}$. It follows that $c_\varphi(nh) = n\varphi(h) = nc_\varphi(h)$ for all $n$ and for $0 \leqslant t \leqslant 1$, we have by (5) and convexity of $c_\varphi$ that

$$tc_\varphi(h) = \varphi(th) \leqslant c_\varphi(th) \leqslant tc_\varphi(h).$$

It follows that $c_\varphi(th) = tc_\varphi(h) = t\varphi(h)$ for all $t \geqslant 0$. Since also $[\varphi^{-h}] = [[\varphi^h]^{-h}] = \varphi$, we have $c_\varphi(-th) = tc_\varphi(-h) = -t\varphi(h)$ for $t \geqslant 0$.

It is easy to see that $c_\varphi(k - \lambda) = c_\varphi(k) - \lambda$ for all $k \in B_\varphi$ and $\lambda \in \mathbb{R}$. Let $\lambda = \varphi(h)$, then it follows that

$$c_\varphi\big(t(h - \lambda)\big) = 0 = c_\varphi\big(t(-h + \lambda)\big)$$

for all $t \geqslant 0$. This implies $\|h - \lambda\|_\varphi = 0$ and hence $h = \lambda$.

Let now $[\varphi^h] = [\varphi^k]$, then $[[\varphi^k]^{-h}] = [\varphi^{k-h}] = \varphi$ and $h - k = \lambda = \varphi(h - k)$. □

Note that the function $\bar{c}_\varphi : B_{\varphi,0} \to \mathbb{R}$ corresponds to the cumulant generating functional in the commutative case. Let us list some of its properties.

**Theorem 11.** *The function $\bar{c}_\varphi$ has the following properties*:

(i) *$\bar{c}_\varphi$ is positive, strictly convex and continuous, $\bar{c}_\varphi(0) = 0$.*
(ii) *$\bar{c}_\varphi$ is Gateaux differentiable, with $\bar{c}'_\varphi(h) = [\varphi^h] - \varphi$.*
(iii) *The map*

$$B_{\varphi,0} \ni h \mapsto \big[\varphi^h\big] - \varphi \in B^*_{\varphi,0}$$

*is one-to-one and norm to $\sigma(B^*_{\varphi,0}, B_{\varphi,0})$-continuous.*

**Proof.** (i) By Corollary 7.1, $\bar{c}_\varphi(h) = S(\varphi, [\varphi^h]) \geqslant 0$ and $\bar{c}_\varphi(0) = 0$ by definition. Let now $h, k \in B_{\varphi,0}$ and $0 < \lambda < 1$ be such that

$$\bar{c}_\varphi\big(\lambda h + (1-\lambda)k\big) = \lambda \bar{c}_\varphi(h) + (1-\lambda)\bar{c}_\varphi(k).$$

Then

$$\begin{aligned}&\sup_{\mathcal{S}_\varphi} \lambda\big(\omega(h) - S(\omega,\varphi)\big) + (1-\lambda)\big(\omega(k) - S(\omega,\varphi)\big)\\ &\quad= \lambda \sup_{\mathcal{S}_\varphi}\big(\omega(h) - S(\omega,\varphi)\big) + (1-\lambda)\sup_{\mathcal{S}_\varphi}\big(\omega(k) - S(\omega,\varphi)\big).\end{aligned}$$

This implies that the maximum in both expressions on the right-hand side is attained at the same point. Therefore $[\varphi^h] = [\varphi^k]$, hence $h - k = \varphi(h-k) = 0$.

(ii) By Theorem 4, $[\varphi^h] - \varphi$ is the unique element in $B^*_{\varphi,0}$, such that

$$\big([\varphi^h] - \varphi\big)(h) = \bar{c}_\varphi(h) + \bar{c}^*_\varphi\big([\varphi^h] - \varphi\big).$$

By [6], this implies that $\bar{c}_\varphi$ is Gateaux differentiable in $h$ with derivative $\bar{c}'_\varphi(h) = [\varphi^h] - \varphi$.

(iii) Let $h_n \to h$ in $B_\varphi$, then $[\varphi^{h_n}]$ converges strongly to $[\varphi^h]$ and $S([\varphi^{h_n}], \varphi) \to S([\varphi^h], \varphi)$. It follows that $[\varphi^{h_n}](k) \to [\varphi^h](k)$ for each $k \in \mathcal{M}_s$ and moreover, the set $\{[\varphi^{h_n}], n \in \mathbb{N}\}$ is equicontinuous in $B^*_\varphi$. This implies that $[\varphi^{h_n}](k) \to [\varphi^h](k)$ for all $k \in B_\varphi$. The map is one-to-one by Corollary 7.2. □

## 8. A manifold structure on faithful states

Recall that a $C^p$-atlas on a set $X$ is a family of pairs $\{(U_i, e_i)\}$, such that

(i) $U_i \subset X$ for all $i$ and $\bigcup U_i = X$.
(ii) For all $i$, $e_i$ is a bijection of $U_i$ onto an open subset $e_i(U_i)$ in some Banach space $B_i$, and for $i, j$, $e_i(U_i \cap U_j)$ is open in $B_i$.
(iii) The map $e_j e_i^{-1}$: $e_i(U_i \cap U_j) \to e_j(U_i \cap U_j)$ is a $C^p$-isomorphism for all $i, j$.

Let $\mathcal{F}_*$ be the set of faithful normal states on $\mathcal{M}$. For $\varphi \in \mathcal{F}_*$, let $V_\varphi$ be the open unit ball in $B_{\varphi,0}$ and let $s_\varphi : V_\varphi \to \mathcal{F}_*$ be the map $h \mapsto [\varphi^h]$. By Corollary 7.2, $s_\varphi$ is a bijection onto the set $s_\varphi(V_\varphi) =: U_\varphi \subset \mathcal{S}_\varphi$. Let $e_\varphi$ be the restriction of $s_\varphi^{-1}$ to $U_\varphi$. Then we have

**Theorem 12.** *$\{(U_\varphi, e_\varphi), \varphi \in \mathcal{F}_*\}$ is a $C^\infty$-atlas on $\mathcal{F}_*$.*

**Proof.** The property (i) and the first part of (ii) of the definition of the $C^p$ atlas are obviously satisfied. Let $\varphi_1, \varphi_2 \in \mathcal{F}_*$ be such that $U_{\varphi_1} \cap U_{\varphi_2} \neq \emptyset$. We prove that $e_{\varphi_1}(U_{\varphi_1} \cap U_{\varphi_2})$ is open in $B_{\varphi_1,0}$.

Let $h_1 \in e_{\varphi_1}(U_{\varphi_1} \cap U_{\varphi_2})$. Then there is some $h_2 \in B_{\varphi_2,0}$, such that $[\varphi_1^{h_1}] = [\varphi_2^{h_2}]$. By Theorem 9, $B_{\varphi_1} = B_{[\varphi_1^{h_1}]} = B_{[\varphi_2^{h_2}]} = B_{\varphi_2}$ and by the chain rule, $\varphi_1 = [\varphi_2^k]$, where $k = h_2 - h_1 + \varphi_2(h_1) \in B_{\varphi_2,0}$. Clearly, the map $B_{\varphi_1,0} \to B_{\varphi_2,0}$, given by $h \mapsto h - \varphi_2(h)$ is continuous.

Let $\varepsilon > 0$ be such that $h_2 + h_2' \in V_{\varphi_2}$ whenever $\|h_2'\|_{\varphi_2} < \varepsilon$ and let us choose $\delta > 0$ such that $h_1 + h_1' \in V_{\varphi_1}$ and $\|h_1' - \varphi_2(h_1')\|_{\varphi_2} < \varepsilon$ for $\|h_1'\|_{\varphi_1} < \delta$. For such $h_1'$, we have

$$s_{\varphi_1}(h_1 + h_1') = \left[\varphi_1^{h_1+h_1'}\right] = \left[\varphi_2^{k+h_1+h_1'-\varphi_2(h_1')}\right] = \left[\varphi_2^{h_2+h_1'-\varphi_2(h_1')}\right] \in U_{\varphi_1} \cap U_{\varphi_2}.$$

This proves that $s_{\varphi_1}^{-1}(U_{\varphi_1} \cap U_{\varphi_2})$ is open in $B_{\varphi_1,0}$. It is also clear that the map

$$\begin{aligned} s_{\varphi_2}^{-1} s_{\varphi_1} : s_{\varphi_1}^{-1}(U_{\varphi_1} \cap U_{\varphi_2}) &\to s_{\varphi_2}^{-1}(U_{\varphi_1} \cap U_{\varphi_2}) \\ h &\mapsto k + h - \varphi_2(h) \end{aligned}$$

is $C^\infty$, which proves (iii). □

It is not difficult to see that for $\varphi \in \mathcal{F}_*$, the set $\mathcal{F}_\varphi := \{[\varphi^h],\ h \in B_{\varphi,0}\}$ is a connected component of the manifold. Let us now define a family of mappings

$$U^{(e)}_{\varphi_1,\varphi_2} : B_{\varphi_1,0} \ni h \mapsto h - \varphi_2(h) \in B_{\varphi_2,0}, \quad \varphi_1, \varphi_2 \in \mathcal{F}_\varphi.$$

It is clear that this defines a parallel transport on the tangent bundle of $\mathcal{F}_\varphi$ and the associated globally flat affine connection is the exponential connection [7].

Let us recall that the dual connection is defined on the cotangent bundle $T^*\mathcal{F}_\varphi$ by means of the parallel transport $\{(U^{(e)}_{\varphi_2,\varphi_1})^*,\ \varphi_1, \varphi_2 \in \mathcal{F}_\varphi\}$, where

$$\left\langle \left(U^{(e)}_{\varphi_2,\varphi_1}\right)^* v, h \right\rangle = \left\langle v, U^{(e)}_{\varphi_1,\varphi_2} h \right\rangle, \quad v \in B^*_{\varphi_2,0},\ h \in B_{\varphi_1,0},$$

and the duality is given by $\langle v, h\rangle = v(h)$. Since $v(h - \varphi_1(h)) = v(h)$ for all $\varphi_1$, the dual parallel transport is

$$U^{(m)}_{\varphi_1,\varphi_2} : B^*_{\varphi_1,0} \ni v \mapsto v \in B^*_{\varphi_2,0}, \quad \varphi_1, \varphi_2 \in \mathcal{F}_\varphi,$$

which corresponds to the mixture connection.

## References


[1] M.H. Al-Rashed, B. Zegarliński, Noncommutative Orlicz spaces associated to a state, Studia Math., in press.

[2] S. Amari, Differential-Geometrical Methods in Statistic, Lecture Notes in Statist., vol. 28, Springer, New York, 1985.

[3] S. Amari, H. Nagaoka, Methods of Information Geometry, Amer. Math. Soc., Providence, RI, 2000.

[4] S. Amari, O.E. Barndorff-Nielsen, R.E. Kass, S.L. Lauritzen, C.R. Rao, Differential Geometry in Statistical Inference, IMS Lecture Notes Monogr. Ser., vol. 10, Inst. Math. Statist., Mayward, 1987.

[5] B. Effron, Defining the curvature of a statistical problem, Ann. Statist. 3 (1975) 1189–1242.

[6] I. Ekeland, R. Temam, Convex Analysis and Variational Problems, North-Holland, Amsterdam, 1976.

[7] P. Gibilisco, G. Pistone, Connections on nonparametric statistical manifolds by Orlicz space geometry, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 1 (1998) 325–347.

[8] M.R. Grasselli, Dual connections in nonparametric classical information geometry, Ann. Inst. Statist. Math., in press, math-ph/0104031.

[9] M. Grasselli, R.F. Streater, The quantum information manifold for epsilon-bounded forms, Rep. Math. Phys. 46 (2000) 325–335, math-ph/9910031.

[10] H. Hasegawa, D. Petz, Noncommutative extension of information geometry II, in: O. Hirota, et al. (Eds.), Quantum Communication, Computing and Measurement, Plenum Press, New York, 1997.

[11] H. Jeffreys, An invariant form of the prior probability in estimation problems, Proc. Roy. Soc. London Ser. A 186 (1946) 453–461.
[12] A. Jenčová, Geometry of quantum states: Dual connections and divergence functions, Rep. Math. Phys. 47 (2001) 121–138.
[13] A. Jenčová, Quantum information geometry and noncommutative $L_p$ spaces, Infin. Dimens. Anal. Quantum Probab. Relat. Top. 8 (2005) 215–233.
[14] G. Köthe, Topological Vector Spaces 1, Springer, Berlin, 1983.
[15] H. Nagaoka, Differential geometrical aspects of quantum state estimation and relative entropy, in: O. Hirota, et al. (Eds.), Quantum Communication, Computing and Measurement, Plenum Press, New York, 1994.
[16] M. Ohya, D. Petz, Quantum Entropy and its Use, Springer, Berlin, 1993.
[17] D. Petz, Information geometry of quantum states, in: R.L. Hudson, et al. (Eds.), Quantum Probability Communications, vol. 10, World Scientific, 1998, pp. 135–158.
[18] G. Pistone, M.P. Rogantin, The exponential statistical manifold, mean parameters, orthogonality and space transformations, Bernoulli 5 (1999) 721–760.
[19] G. Pistone, C. Sempi, An infinite-dimensional geometric structure on the space of all probability measures equivalent to a given one, Ann. Statist. 23 (1995) 1543–1561.
[20] C.R. Rao, Information and accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. Soc. 37 (1945) 81–91.
[21] R.F. Streater, The information manifold for relatively bounded potentials, Tr. Mat. Inst. Steklova 228 (2000) 217–235.
[22] R.F. Streater, The analytic quantum information manifold, in: Stochastic Processes, Physics and Geometry: New Interplays, II, Leipzig, 1999, Amer. Math. Soc., Providence, RI, 2000, pp. 603–611, math-ph/9910036.
[23] R.F. Streater, Quantum Orlicz spaces in information geometry, Open Sys. Inf. Dyn. 11 (2004) 359–375, math-ph/0407046.

# 16
# On quantum information manifolds

Anna Jenčová

## 16.1 Introduction

The aim of information geometry is to introduce a suitable geometrical structure on families of probability distributions or quantum states. For parametrised statistical models, such structure is based on two fundamental notions: the Fisher information and the exponential family with its dual mixed parametrisation, see for example (Amari 1985, Amari and Nagaoka 2000).

For the non-parametric situation, the solution was given by Pistone and Sempi (Pistone and Sempi 1995, Pistone and Rogantin 1999), who introduced a Banach manifold structure on the set $\mathcal{P}$ of probability distributions, equivalent to a given one. For each $\mu \in \mathcal{P}$, the authors considered the non-parametric exponential family at $\mu$. As it turned out, this provides a $C^\infty$-atlas on $\mathcal{P}$, with the exponential Orlicz spaces $L_\Phi(\mu)$ as the underlying Banach spaces, here $\Phi$ is the Young function of the form $\Phi(x) = \cosh(x) - 1$.

The present contribution deals with the case of quantum states: we want to introduce a similar manifold structure on the set of faithful normal states of a von Neumann algebra $\mathcal{M}$. Since there is no suitable definition of a non-commutative Orlicz space with respect to a state $\varphi$, it is not clear how to choose the Banach space for the manifold. Of course, there is a natural Banach space structure, inherited from the predual $\mathcal{M}_*$. But, as it was already pointed out in (Streater 2004), this structure is not suitable to define the geometry of states: for example, any neighbourhood of a state $\varphi$ contains states such that the relative entropy with respect to $\varphi$ is infinite.

In (Jenčová 2006), we suggest the following construction. We define a Luxemburg norm using a quantum Young function, similar to that in (Streater 2004) but restricted to the space of self-adjoint operators in $\mathcal{M}$. Then we take the completion under this norm. In the classical case, this norm coincides with the norm of Pistone and Sempi, restricted to bounded measurable functions. This is described in Section 16.2. In Section 16.3, we show that an equivalent Banach space can be obtained in a more natural and easier way, using some results of convex analysis. In the following sections, we use the results in (Jenčová 2006) to introduce the manifold, and discuss possible extensions.

*Algebraic and Geometric Methods in Statistics*, ed. Paolo Gibilisco, Eva Riccomagno, Maria Piera Rogantin and Henry P. Wynn. Published by Cambridge University Press. 

Section 16.6 is devoted to channels, that is, completely positive unital maps between the algebras. We show that the structures we introduced are closely related to sufficiency of channels and a new characterisation of sufficiency is given. As it turns out, the new definition of the spaces provides a convenient way to deal with these problems.

## 16.2 The quantum Orlicz space

We recall the definition and some properties of the quantum exponential Orlicz space, as given in (Jenčová 2006).

### 16.2.1 Young functions and associated norms

Let $V$ be a real Banach space and let $V^*$ be its dual. We say that a function $\Phi : V \to \mathbb{R} \cup \{\infty\}$ is a *Young function*, if it satisfies:

(i) $\Phi$ is convex and lower semicontinuous;
(ii) $\Phi(x) \geq 0$ for all $x \in V$ and $\Phi(0) = 0$,
(iii) $\Phi(x) = \Phi(-x)$ for all $x \in V$,
(iv) if $x \neq 0$, then $\lim_{t\to\infty} \Phi(tx) = \infty$.

Since $\Phi$ is convex, its effective domain

$$\mathrm{dom}(\Phi) := \{x \in V,\ \Phi(x) < \infty\}$$

is a convex set. Let us define the sets

$$\begin{aligned} C_\Phi &:= \{x \in V, \Phi(x) \leq 1\}, \\ L_\Phi &:= \{x \in V, \exists s > 0,\ \text{such that } \Phi(sx) < \infty\}. \end{aligned}$$

Then $L_\Phi$ is the smallest vector space, containing $\mathrm{dom}(\Phi)$. Moreover, the Minkowski functional of $C_\Phi$,

$$\|x\|_\Phi := \inf\{\rho > 0, x \in \rho C_\Phi\} = \inf\{\rho > 0, \Phi(\rho^{-1}x) \leq 1\}$$

defines a norm in $L_\Phi$.

Let $B_\Phi$ be the completion of $L_\Phi$ under $\|\cdot\|_\Phi$. If the function $\Phi$ is finite valued, $\Phi : V \to \mathbb{R}$, (or, more generally, $0 \in \mathrm{int\ dom}(\Phi)$), then $L_\phi = V$ and the norm $\|\cdot\|_\Phi$ is continuous with respect to the original norm in $V$, so that we have the continuous inclusion $V \sqsubseteq B_\Phi$.

Let now $\Phi : V \to \mathbb{R}$ be a Young function and let the function $\Phi^* : V^* \to \mathbb{R} \cup \{\infty\}$ be the conjugate of $\Phi$,

$$\Phi^*(v) = \sup_{x\in V} v(x) - \Phi(x)$$

then $\Phi^*$ is a Young function as well. The associated norm satisfies

$$|v(x)| \leq 2\|x\|_\Phi \|v\|_{\Phi^*} \qquad x \in B_\Phi,\ v \in B_{\Phi^*}$$

(the Hölder inequality), so that each $v \in B_{\Phi^*}$ defines a continuous linear functional on $B_\Phi$, in fact, it can be shown that

$$L_{\Phi^*} = B_{\Phi^*} = B_\Phi^* \sqsubseteq V^*$$

in the sense that the norm $\|\cdot\|_{\phi^*}$ is equivalent with the usual norm in $B^*_\Phi$. Similarly, we have $L_\Phi = V \sqsubseteq B_\Phi \subseteq B^*_{\Phi^*}$.

### 16.2.2 Relative entropy

Let $\mathcal{M}$ be a von Neumann algebra in standard form. Let $\mathcal{M}^+_*$ be the set of normal positive linear functionals and $\mathfrak{S}_*$ be the set of normal states on $\mathcal{M}$. For $\omega$ and $\varphi$ in $\mathcal{M}^+_*$, the relative entropy is defined as

$$S(\omega,\varphi) = \begin{cases} -\langle \log(\Delta_{\varphi,\xi_\omega})\xi_\omega, \xi_\omega\rangle & \text{if } \operatorname{supp}\omega \le \operatorname{supp}\varphi \\ \\ \infty & \text{otherwise} \end{cases}$$

where $\xi_\omega$ is the representing vector of $\omega$ in a natural positive cone and $\Delta_{\varphi,\xi_\omega}$ is the relative modular operator. Then $S$ is jointly convex and weakly lower semicontinuous. We will also need the following identity

$$S(\psi_\lambda,\varphi) + \lambda S(\psi_1,\psi_\lambda) + (1-\lambda)S(\psi_2,\psi_\lambda) = \lambda S(\psi_1,\varphi) + (1-\lambda)S(\psi_2,\varphi) \quad (16.1)$$

where $\psi_1$, $\psi_2$ are normal states and $\psi_\lambda = \lambda\psi_1 + (1-\lambda)\psi_2$, $0 \le \lambda \le 1$. This implies that $S$ is strictly convex in the first variable.

Let us denote

$$\begin{aligned} \mathcal{P}_\varphi &:= \{\omega \in \mathcal{M}^+_*, S(\omega,\varphi) < \infty\} \\ \mathcal{S}_\varphi &:= \{\omega \in \mathfrak{S}_*,\ S(\omega,\varphi) < \infty\} \\ K_{\varphi,C} &:= \{\omega \in \mathfrak{S}_*,\ S(\omega,\varphi) \le C\}, \qquad C > 0. \end{aligned}$$

Then $\mathcal{P}_\varphi$ is a convex cone dense in $\mathcal{M}^+_*$ and $\mathcal{S}_\varphi$ is a convex set generating $\mathcal{P}_\varphi$. By (16.1), $\mathcal{S}_\varphi$ is a face in $\mathfrak{S}_*$. For any $C > 0$, the set $K_{\varphi,C}$ separates the elements in $\mathcal{M}$ and it is convex and compact in the $\sigma(\mathcal{M}_*,\mathcal{M})$-topology.

### 16.2.3 The quantum exponential Orlicz space and its dual

Let $\mathcal{M}_s$ be the real Banach subspace of self-adjoint elements in $\mathcal{M}$, then the dual $\mathcal{M}^*_s$ is the subspace of Hermitian (not necessarily normal) functionals in $\mathcal{M}^*$. We define the functional $F_\varphi :\ \mathcal{M}^*_s \to \mathbb{R} \cup \{\infty\}$ by

$$F_\varphi(\omega) = \begin{cases} S(\omega,\varphi) & \text{if } \omega \in \mathfrak{S}_* \\ \infty & \text{otherwise.} \end{cases}$$

Then $F_\varphi$ is strictly convex and lower semicontinuous; with $\operatorname{dom}(F_\varphi) = \mathcal{S}_\varphi$. Its conjugate

$$F^*_\varphi(h) = \sup_{\omega\in\mathfrak{S}_*} \omega(h) - S(\omega,\varphi)$$

is convex and lower semicontinuous; in fact, being finite valued, it is continuous on $\mathcal{M}_s$. We have $F^{**}_\varphi = F_\varphi$ on $\mathcal{M}^*_s$.

We define the function $\Phi_\varphi : \mathcal{M}_s \to \mathbb{R}$ by

$$\Phi_\varphi(h) = \frac{\exp(F^*_\varphi(h)) + \exp(F^*_\varphi(-h))}{2} - 1.$$

Then $\Phi_\varphi$ is a Young function. Let us denote $\|h\|_\varphi := \|h\|_{\Phi_\varphi}$ and $B_\varphi := B_{\Phi_\varphi}$, then we call $B_\varphi$ the *quantum exponential Orlicz space.*

Let $h \in \mathcal{M}_s$, $\|h\|_\varphi \leq 1$. Then

$$\cosh(\omega(h)) \leq 2e^{S(\omega,\varphi)}.$$

It follows that each $\omega \in \mathcal{S}_\varphi$ defines a continuous linear functional on $B_\varphi$. We denote by $B_{\varphi,0}$ the Banach subspace of centred elements in $B_\varphi$, that is, $h \in B_\varphi$ with $\varphi(h) = 0$. Then

$$\Phi_{\varphi,0}(h) = \frac{F_\varphi^*(h) + F_\varphi^*(-h)}{2}$$

is a Young function on $\mathcal{M}_{s,0} := \{h \in \mathcal{M}_s, \varphi(h) = 0\}$ and it defines an equivalent norm in $B_{\varphi,0}$.

**Remark 16.1** Let $\mathcal{M}$ be commutative, then $\mathcal{M} = L_\infty(X, \Sigma, \mu)$ for some measure space $(X, \Sigma, \mu)$ with $\sigma$-finite measure $\mu$. Then $\varphi$ is a probability measure on $\Sigma$, with the density $p := d\varphi/d\mu \in L_1(X, \Sigma, \mu)$. For any Hermitian element $u \in \mathcal{M}$, $F_\varphi^*(u) = \log \int \exp(u) p d\mu$, so that

$$\Phi_\varphi(u) = \int \cosh(u) p d\mu - 1.$$

It follows that in this case, our space $B_\varphi$ coincides with the closure $M_\Phi(\varphi)$ of $L_\infty(X, \Sigma, \varphi)$ in $L_\Phi(\varphi)$.

Let us now describe the dual space $B_{\varphi,0}^*$. It was proved that $B_\varphi^* = \mathcal{P}_\varphi - \mathcal{P}_\varphi$ and $B_{\varphi,0}^* = \cup_n n(K_{\varphi,1} - K_{\varphi,1})$. If we denote by $C_{\varphi,0}$ the closed unit ball in $B_{\varphi,0}^*$, then

$$C_{\varphi,0} \subseteq K_{\varphi,1} - K_{\varphi,1} \subseteq 4C_{\varphi,0} \tag{16.2}$$

so that any element in $C_{\varphi,0}$ can be written as a difference of two states in $K_{\varphi,1}$.

For $v$ in $\mathcal{S}_\varphi - \mathcal{S}_\varphi$, let $L_v := \{\omega_1, \omega_2 \in \mathcal{S}_\varphi,\ v = \omega_1 - \omega_2\}$. We define the function $\Psi_{\varphi,0} :\ \mathcal{M}_{s,0}^* \to \mathbb{R}^+$ by

$$\Psi_{\varphi,0}(v) = \begin{cases} \inf_{L_v} S(\omega_1, \varphi) + S(\omega_2, \varphi) & \text{if } v \in \mathcal{S}_\varphi - \mathcal{S}_\varphi \\ \\ \infty & \text{otherwise.} \end{cases}$$

Then $\Psi_{\varphi,0}$ is a Young function and it was proved that

$$\Phi_{\varphi,0}^*(v) = 1/2\Psi_{\varphi,0}(2v)$$

for $v \in \mathcal{M}_{s,0}^*$. It follows that the norm in $B_{\varphi,0}^*$ is equivalent with $\|\cdot\|_{\Psi_{\varphi,0}}$.

### 16.3 The spaces $A(K_\varphi)$ and $A(K_\varphi)^{**}$

In this section, we use a well-known representation of compact convex sets, see for example (Asimow and Ellis 1980) for details. We obtain a Banach space, which turns out to be equivalent to $B_{\varphi,0}$.

Let $K \subset \mathfrak{S}_*$ be a convex set, compact in the $\sigma(\mathcal{M}_*, \mathcal{M})$-topology and separating the points in $\mathcal{M}_s$. In particular, let $K_\varphi := K_{\varphi,1}$. Let $A(K)$ be the Banach space of continuous affine functions $f : K \to \mathbb{R}$, with the supremum norm. Then $K$ can be identified with the set of states on $A(K)$, where each element $\omega \in K$ acts on $A(K)$ by evaluation $f \mapsto f(\omega)$. Moreover, the topology of $K$ coincides with the weak*-topology of the state space.

It is clear that any self-adjoint element in $\mathcal{M}$ belongs to $A(K)$, moreover, $\mathcal{M}_s$ is a linear subspace in $A(K)$, separating the points in $K$ and containing all the constant functions. It follows that $\mathcal{M}_s$ is norm-dense in $A(K)$.

The dual space $A(K)^*$ is the set of all elements of the form

$$p : f \mapsto a_1 f(\omega_1) - a_2 f(\omega_2)$$

for some $\omega_1, \omega_2 \in K$, $a_1, a_2 \in \mathbb{R}^+$, so that $A(K)^*$ is a real linear subspace in $\mathcal{M}_*$. The embedding of $A(K)^*$ to $\mathcal{M}_*$ is continuous and the weak*-topology on $A(K)^*$ coincides with $\sigma(\mathcal{M}_*, \mathcal{M})$ on bounded subsets. It is also easy to see that the second dual $A(K)^{**}$ is the set of all bounded affine functionals on $K$.

Let $L \subseteq K$ be convex and compact. For $f \in A(K)^{**}$, the restriction to $L$ is in $A(L)^{**}$, continuous if $f \in A(K)$ and such that $\|f|_L\|_L \leq \|f\|_K$.

**Lemma 16.1** *Let $a, b > 0$, then $A(K_{\varphi,a}) = A(K_{\varphi,b})$ and $A(K_{\varphi,a})^{**} = A(K_{\varphi,b})^{**}$, in the sense that the corresponding norms are equivalent.*

*Proof* Suppose that $a \geq b$. Since $K_{\varphi,b} \subseteq K_{\varphi,a}$, it follows that $A(K_{\varphi,a}) \subseteq A(K_{\varphi,b})$ and $A(K_{\varphi,a})^{**} \subseteq A(K_{\varphi,b})^{**}$ with $\|f\|_{\varphi,b} \leq \|f\|_{\varphi,a}$ for $f \in A(K_{\varphi,a})^{**}$. On the other hand, let $\omega \in K_{\varphi,a}$, then $\omega_t := t\omega + (1-t)\varphi \in K_{\varphi,b}$ whenever $t \leq b/a$. Then

$$\omega = a/b\omega_{b/a} - (a/b - 1)\varphi$$

so that $K_{\varphi,a}$ is contained in the closed ball with radius $(2a-b)/b$ in $A(K_{\varphi,b})^*$. It follows that any $f \in A(K_{\varphi,b})^{**}$ defines a bounded affine functional over $K_{\varphi,a}$, continuous if $f \in A(K_{\varphi,b})$ and

$$\|f\|_{\varphi,a} = \sup_{\omega \in K_{\varphi,a}} |f(\omega)| \leq \|f\|_{\varphi,b}(2a-b)/b.$$

□

We see from the above proof that $\mathcal{S}_\varphi \subset A(K_{\varphi,b})^*$ and each $K_{\varphi,a}$ is weak*-compact in $A(K_{\varphi,b})^*$. It follows that $\operatorname{dom}(F_\varphi) \subset A(K_{\varphi,b})^*$ and $F_\varphi$ is a convex weak*-lower semicontinuous functional on $A(K_{\varphi,b})^*$.

Let us denote by $A_0(K)$ the subspace of elements $f \in A(K)$, such that $f(\varphi) = 0$. Then we have

**Theorem 16.1** $A_0(K_\varphi) = B_{\varphi,0}$, *with equivalent norms.*

*Proof* We have by (16.2) that the norms are equivalent on $\mathcal{M}_s$. The statement follows from the fact that $\mathcal{M}_s$ is dense in both spaces. □

### 16.4 The perturbed states

As we have seen, $F_\varphi(\omega) = S(\omega, \varphi)$ defines a convex lower semicontinuous functional $A(K_\varphi)^* \to \mathbb{R}$. Let $f \in A(K_\varphi)^{**}$. We denote

$$c_\varphi(f) := \inf_{\omega \in \mathcal{S}_\varphi} f(\omega) + S(\omega, \varphi).$$

Then $-c_\varphi(-f)$ is the conjugate functional $F_\varphi^*(f)$, so that $c_\varphi$ is concave and upper semicontinuous, with values in $\mathbb{R} \cup \{-\infty\}$.

Suppose that $c_\varphi(f)$ is finite and that there is a state $\psi \in \mathcal{S}_\varphi$, such that

$$c_\varphi(f) = f(\psi) + S(\psi, \varphi).$$

Then this state is unique, this follows from the fact that $S$ is strictly convex in the first variable. Let us denote this state by $\varphi^f$. Note that if $f \in \mathcal{M}_s$, then $\varphi^f$ exists and it is the perturbed state (Ohya and Petz 1993), so that we can see the mapping $f \mapsto \varphi^f$ as an extension of state perturbation.

In (Jenčová 2006), we defined the perturbed state for elements in $B_{\varphi,0}$; we remark that there we used the notation $c_\varphi = F_\varphi^*$ and the state was denoted by $[\varphi^h]$, $h \in B_\varphi$. It was shown that $[\varphi^h]$ is defined for all $h \in B_\varphi$ and that the map

$$B_{\varphi,0} \in h \mapsto [\varphi^h]$$

can be used to define a $C^\infty$-atlas on the set of faithful states on $\mathcal{M}$. By Theorem 16.1, we have the same for $A_0(K_\varphi)$. We will recall the construction below, but before that, we give some results obtained for $f \in A(K_\varphi)^{**}$.

First of all, it is clear that $c_\varphi(f + c) = c_\varphi(f) + c$ for any real $c$ and $\varphi^f = \varphi^{f+c}$ if $\varphi^f$ is defined. We may therefore suppose that $f \in A_0(K_\varphi)^{**}$.

**Lemma 16.2** *Let $f \in A(K_\varphi)^{**}$ be such that $\varphi^f$ exists. Then for all $\omega \in \mathcal{S}_\varphi$,*

$$S(\omega, \varphi) + f(\omega) \geq S(\omega, \varphi^f) + c_\varphi(f).$$

*Equality is attained on the face in $\mathcal{S}_\varphi$, generated by $\varphi^f$.*

*Proof* The statement is proved using the identity (16.1), the same way as Lemmas 12.1 and 12.2 in (Ohya and Petz 1993). □

The previous lemma has several consequences. For example, it follows that $c_\varphi(f) \leq -S(\varphi, \varphi^f) \leq 0$ if $f \in A_0(K_\varphi)^{**}$. Further, $S(\omega, \varphi^f)$ is bounded on $K_\varphi$, so that $K_\varphi \subseteq K_{\varphi^f, C}$ for some $C > 0$. It also follows that $\mathcal{S}_\varphi \subseteq \mathcal{S}_{\varphi^f}$. In particular, $S(\varphi, \varphi^f) < \infty$ and since also $S(\varphi^f, \varphi) < \infty$, the states $\varphi$ and $\varphi^f$ have the same support.

**Lemma 16.3** *Let $\psi = \varphi^f$ for some $f \in A(K_\varphi)^{**}$. Then we have the continuous embeddings $A(K_\psi) \sqsubseteq A(K_\varphi)$ and $A(K_\psi)^{**} \sqsubseteq A(K_\varphi)^{**}$.*

*Proof* Follows from $K_\varphi \subseteq K_{\psi,C}$ and Lemma 16.1. □

We will now consider the set of all states $\varphi^f$, with some $f \in A(K_\varphi)^{**}$. Let $\psi$ be a normal state, such that $\varphi \in \mathcal{S}_\psi$. We denote

$$f_\psi(\omega) := S(\omega, \psi) - S(\omega, \varphi) - S(\varphi, \psi).$$

By identity (16.1), $f_\psi$ is an affine functional $K_\varphi \to \mathbb{R} \cup \{\infty\}$, such that $f_\psi(\varphi) = 0$.

**Theorem 16.2** *Let $\psi$ be a normal state. Then $\psi = \varphi^f$ for some $f \in A(K_\varphi)^{**}$ if and only if $\psi \in \mathcal{S}_\varphi$ and $K_\varphi \subseteq K_{\psi,C}$ for some $C > 0$.*

*Proof* It is clear that $\psi \in \mathcal{S}_\varphi$ if $\psi = \varphi^f$ and we have seen that also $K_\varphi \subseteq K_{\psi,C}$. Conversely, if $K_\varphi \subseteq K_{\psi,C}$, then $f_\psi \in A_0(K_\varphi)^{**}$ and

$$f_\psi(\omega) + S(\omega, \varphi) = S(\omega, \psi) - S(\varphi, \psi) \geq -S(\varphi, \psi)$$

for all $\omega \in \mathcal{S}_\varphi$. Since equality is attained for $\omega = \psi$, $\psi = \varphi^{f_\psi}$. $\square$

Note also that, by the above proof, $c_\varphi(f_\psi) = -S(\varphi, \psi)$.

### *16.4.1 The subdifferential*

Let $\psi \in \mathcal{S}_\varphi$. The *subdifferential at* $\psi$ is the set of elements $f \in A_0(K_\varphi)^{**}$, such that $\psi = \varphi^f$. Let us denote the subdifferential by $\partial_\varphi(\psi)$. By Theorem 16.2, the subdifferential at $\psi$ is non-empty if and only if $K_\varphi \subseteq K_{\psi,C}$.

**Lemma 16.4** *If $\partial_\varphi(\psi) \neq \emptyset$, then it is a closed convex subset in $A_0(K_\varphi)^{**}$. Moreover, $c_\varphi$ is affine over $\partial_\varphi(\psi)$.*

*Proof* Let $f, g \in \partial_\varphi(\psi)$ and let $g_\lambda = \lambda g + (1-\lambda) f$, $\lambda \in (0,1)$. Then

$$g_\lambda(\psi) + S(\psi, \varphi) = \lambda c_\varphi(g) + (1-\lambda) c_\varphi(f).$$

Since $c_\varphi$ is concave, this implies that $\psi = \varphi^{g_\lambda}$ and that $c_\varphi(g_\lambda) = \lambda c_\varphi(g) + (1 - \lambda) c_\varphi(f)$. Moreover, we can write

$$\partial_\varphi(\psi) = \{g \in A(K_\varphi)^{**},\ c_\varphi(g) - g(\psi) \geq S(\psi, \varphi)\}$$

and this set is closed, since $c_\varphi$ is upper semicontinuous. $\square$

**Lemma 16.5** *Let $\psi \in \mathcal{S}_\varphi$, $\partial_\varphi(\psi) \neq \emptyset$ and let $g \in A_0(K_\varphi)^{**}$. Then $g \in \partial_\varphi(\psi)$ if and only if there is some $k \in \mathbb{R}$, such that*

$$g(\omega) - f_\psi(\omega) \geq k,\ \omega \in \mathcal{S}_\varphi \quad \textit{and} \quad g(\psi) - f(\psi) = k. \tag{16.3}$$

*In this case, $k = c_\varphi(g) - c_\varphi(f_\psi) \leq 0$.*

*Proof* If $g \in \partial_\varphi(\psi)$, then (16.3) follows from Lemma 16.2 and $k \leq 0$ is obtained by putting $\omega = \varphi$. Conversely, suppose that (16.3) is true, then we have for $\omega \in \mathcal{S}_\varphi$

$$g(\omega) + S(\omega, \varphi) = g(\omega) - f_\psi(\omega) + f_\psi(\omega) + S(\omega, \varphi) \geq k + c_\varphi(f_\psi)$$

and $g(\psi)+S(\psi,\omega)=k+c_\varphi(f_\psi)$, this implies that $\psi=\varphi^g$ and $c_\varphi(g)=k+c_\varphi(f_\psi)$. □

### *16.4.2 The chain rule*

Let $\mathcal{C}_\varphi := \{\psi \in S_\varphi,\ K_\varphi \subseteq K_{\psi,C},\ K_\psi \subseteq K_{\varphi,A}$ for some $A, C > 0\}$.

**Theorem 16.3** *Let $\psi \in \mathcal{C}_\varphi$. Then*

(i) $\mathcal{S}_\varphi = \mathcal{S}_\psi$,
(ii) $A(K_\varphi) = A(K_\psi)$, $A(K_\varphi)^{**} = A(K_\psi)^{**}$, *with equivalent norms,*
(iii) $\varphi \in \mathcal{C}_\psi$.

*Proof* Let $\psi \in \mathcal{C}_\varphi$. By Theorem 16.2, $\psi = \varphi^f$, $f \in A(K_\varphi)^{**}$ and also $\varphi = \psi^g$ for some $g \in A(K_\psi)^{**}$. Now we have (i) by Lemma 16.2 and (ii) by Lemma 16.3, (iii) is obvious. □

We also have the following chain rule.

**Theorem 16.4** *Let $\psi \in \mathcal{C}_\varphi$ and let $g \in A(K_\varphi)^{**}$ be such that $\psi^g$ exists. Then*

$$c_\psi(g) = c_\varphi(g+f) - c_\varphi(f), \qquad \psi^g = \varphi^{f+g} \tag{16.4}$$

*holds for $f = f_\psi$.*

*Proof* Suppose that $\psi^g$ exists, then

$$g(\omega) + f_\psi(\omega) + S(\omega,\varphi) = g(\omega) + S(\omega,\psi) + c_\varphi(f_\psi) \geq c_\psi(g) + c_\varphi(f_\psi)$$

for all $\omega \in \mathcal{S}_\varphi = \mathcal{S}_\psi$ and equality is attained at $\omega = \psi^g$. This implies $c_\psi(g) = c_\varphi(g+f) - c_\varphi(f)$ and $\psi^g = \varphi^{f+g}$. □

## 16.5 The manifold structure

Let $\mathcal{F}$ be the set of faithful normal states on $\mathcal{M}$. Let $\varphi \in \mathcal{F}$. In this section we show that we can use the map $f \mapsto \varphi^f$ to define the manifold structure on $\mathcal{F}$. So far, it is not clear if this map is well-defined or one-to-one on $A_0(K_\varphi)^{**}$. The situation is better if we restrict to $A_0(K_\varphi)$, as Theorem 16.5 shows.

**Theorem 16.5** *Let $f \in A(K_\varphi)$. Then*

(i) *$\varphi^f$ exists and $\varphi^f \in \mathcal{C}_\varphi$.*
(ii) *If $g \in A(K_\varphi)$ is such that $\varphi^g = \varphi^f$, then $f - g = \varphi(f-g)$.*
(iii) *In Lemma 16.2, equality is attained for all $\omega \in \mathcal{S}_\varphi$, in particular,*

$$f - f(\varphi) = f_{\varphi^f}.$$

(iv) *The chain rule (16.4) holds for all $f, g \in A(K_\varphi)$.*

*Proof* We may suppose that $f \in A_0(K_\varphi)$. By the results in (Jenčová 2006) and Theorem 16.1, if $f \in A_0(K_\varphi) = B_{\varphi,0}$, then $\psi = \varphi^f$ exists, $f - f(\psi) \in A_0(K_\psi) = B_{\psi,0}$ and $\varphi = \psi^{-f}$. By Theorem 16.2, $\psi \in \mathcal{C}_\varphi$ and (i) is proved. (ii),(iii) and (iv) were proved in (Jenčová 2006). □

Proposition 16.1 is not needed in our construction. It shows that each $\psi \in \mathcal{C}_\varphi$ is faithful on $A(K_\varphi)$.

**Proposition 16.1** *Let $\psi \in \mathcal{C}_\varphi$ and let $g \in A(K_\varphi)$ be positive. Then $g(\psi) = 0$ implies $g = 0$.*

*Proof* Let $g$ be a positive element in $A(K_\varphi) = A(K_\psi)$, with $g(\psi) = 0$, then by Lemma 16.5, $f_\psi + g \in \partial_\varphi(\psi)$. Since $\psi^g$ exists, we have by the chain rule that $\psi^g = \varphi^{f_\psi + g} = \psi$. Since $g \in A_0(K_\psi)$, $g = 0$. □

Let us recall that a $C^p$-atlas on a set $X$ is a family of pairs $\{(U_i, e_i)\}$, such that

(i) $U_i \subset X$ for all $i$ and $\cup U_i = X$;
(ii) for all $i$, $e_i$ is a bijection of $U_i$ onto an open subset $e_i(U_i)$ in some Banach space $B_i$, and for all $i, j$, $e_i(U_i \cap U_j)$ is open in $B_i$;
(iii) the map $e_j e_i^{-1} : e_i(U_i \cap U_j) \to e_j(U_i \cap U_j)$ is a $C^p$-isomorphism for all $i, j$.

Let now $X = \mathcal{F}$. For $\varphi \in \mathcal{F}$, let $V_\varphi$ be the open unit ball in $A_0(K_\varphi)$ and let $s_\varphi : V_\varphi \to \mathcal{F}$ be the map $f \mapsto \varphi^f$. By Theorem 16.5, $s_\varphi$ is a bijection onto the set $U_\varphi := s_\varphi(V_\varphi)$. Let $e_\varphi$ be the map $U_\varphi \ni \psi \mapsto f_\psi \in V_\varphi$. Then we have

**Theorem 16.6** *(Jenčová 2006) $\{(U_\varphi, e_\varphi), \varphi \in \mathcal{F}\}$ is a $C^\infty$-atlas on $\mathcal{F}$.*

In the commutative case, the space corresponding to $A(K_\varphi)$ is not the exponential Orlicz space $L_\Phi$, but the subspace $M_\Phi$, see Remark 16.1. The corresponding commutative information manifold structure was considered in (Grasselli 2009). It follows from the theory of Orlicz spaces that (under some reasonable conditions on the base measure $\mu$)

$$M_\Phi(\mu)^* = L_{\Phi^*}(\mu), \qquad L_{\Phi^*}(\mu)^* = L_\Phi(\mu).$$

By comparing $A(K_\varphi)$ with these results, it seems that the quantum exponential Orlicz space should be the second dual $A(K_\varphi)^{**}$, rather than $A(K_\varphi)$.

To get the counterpart of the Pistone and Sempi manifold, we would need to extend the map $s_\varphi$ to the unit ball $V_\varphi^{**}$ in $A_0(K_\varphi)^{**}$ and show that it is one-to-one. At present, it is not clear how to prove this. At least, we can prove that $c_\varphi$ is finite on $V_\varphi^{**}$.

**Lemma 16.6** *Let $f \in A_0(K_\varphi)^{**}$, $\|f\| \leq 1$. Then $0 \geq c_\varphi(f) \geq -1$ and the infimum can be taken over $K_\varphi$.*

*Proof* Let $\omega \in \mathcal{S}_\varphi$ be such that $S(\omega, \varphi) > 1$. Since the function $t \mapsto S(\omega_t, \varphi)$ is convex and lower semicontinuous in $(0, 1)$, it is continuous and there is some

$t \in (0,1)$ such that $S(\omega_t, \varphi) = 1$, recall that $\omega_t = t\omega + (1-t)\varphi$. By strict convexity, it follows that $1 = S(\omega_t, \varphi) < tS(\omega, \varphi)$ and $S(\omega, \varphi) > 1/t$. On the other hand, $\omega_t \in K_\varphi$ and therefore $-1 \le f(\omega_t) = tf(\omega)$. It follows that

$$f(\omega) + S(\omega, \varphi) > -1/t + 1/t = 0 = f(\varphi) + S(\varphi, \varphi) \ge c_\varphi(f).$$

From this, $c_\varphi(f) = \inf_{\omega \in K_\varphi} f(\omega) + S(\omega, \varphi) \ge -1$. □

## 16.6 Channels and sufficiency

Let $\mathcal{N}$ be another von Neumann algebra. A *channel* from $\mathcal{N}$ to $\mathcal{M}$ is a completely positive, unital map $\alpha : \mathcal{N} \to \mathcal{M}$. We will also require that a channel is normal, then its dual $\alpha^* : \varphi \mapsto \varphi \circ \alpha$ maps normal states on $\mathcal{M}$ to normal states on $\mathcal{N}$.

An important property of such channels is that the relative entropy is monotone under these maps:

$$S(\omega \circ \alpha, \varphi \circ \alpha) \le S(\omega, \varphi), \qquad \omega, \varphi \in \mathfrak{S}_*.$$

This implies that $\alpha^*$ defines a continuous affine map $K_\varphi \to K_{\varphi\circ\alpha}$. If $f_0 \in A(K_{\varphi\circ\alpha})^{**}$, then composition with $\alpha^*$ defines a bounded affine functional over $K_\varphi$, which we denote by $\alpha(f_0)$. Then $\alpha(f_0)$ is continuous if $f_0 \in A(K_{\varphi\circ\alpha})$ and

$$\|\alpha(f_0)\| = \sup_{\omega \in K_\varphi} |f_0(\omega \circ \alpha)| \le \sup_{\omega_0 \in K_{\varphi\circ\alpha}} |f_0(\omega_0)| = \|f_0\|$$

so that $\alpha$ is a contraction $A(K_{\varphi\circ\alpha})^{**} \to A(K_\varphi)^{**}$ and $A(K_{\varphi\circ\alpha}) \to A(K_\varphi)$.

**Lemma 16.7** *Let $\alpha : \mathcal{N} \to \mathcal{M}$ be a channel and let $g_0 \in A(K_{\varphi\circ\alpha})^{**}$. Then $c_{\varphi\circ\alpha}(g_0) \le c_\varphi(\alpha(g_0))$.*

*Proof* We compute

$$c_\varphi(\alpha(g_0)) = \inf_{\omega \in \mathcal{S}_\varphi} g_0(\omega \circ \alpha) + S(\omega, \varphi) \ge \inf_{\omega \in \mathcal{S}_\varphi} g_0(\omega \circ \alpha) + S(\omega \circ \alpha, \varphi \circ \alpha) \ge c_{\varphi\circ\alpha}(g_0).$$

□

Let $\mathcal{S}$ be a set of states in $\mathfrak{S}_*(\mathcal{M})$. We say that the channel $\alpha : \mathcal{N} \to \mathcal{M}$ is *sufficient for $\mathcal{S}$* if there is a channel $\beta : \mathcal{M} \to \mathcal{N}$, such that

$$\omega \circ \alpha \circ \beta = \omega, \qquad \omega \in \mathcal{S}.$$

This definition of sufficient channels was introduced in (Petz 1986), see also (Jenčová and Petz 2006a), and several characterisations of sufficiency were given. Here we are interested in the following two characterisations. For simplicity, we will assume that the states, as well as the channel, are faithful.

**Theorem 16.7** *(Petz 1986) Let $\psi \in \mathcal{S}_\varphi$. The channel $\alpha$ is sufficient for the pair $\{\psi, \varphi\}$ if and only if $S(\psi, \varphi) = S(\psi \circ \alpha, \varphi \circ \alpha)$.*

**Theorem 16.8** *(Jenčová and Petz 2006b) Let $\psi = \varphi^f$ for some $f \in \mathcal{M}_s$. Then $\alpha$ is sufficient for $\{\psi, \varphi\}$ if and only if there is some $g_0 \in \mathcal{N}_s$, such that $\psi \circ \alpha = (\varphi \circ \alpha)^{g_0}$ and $f = \alpha(g_0)$.*

In this section we show how Theorem 16.8 can be extended to pairs $\{\psi, \varphi\}$ such that $\partial_\varphi(\psi) \neq \emptyset$.

So let $\psi = \varphi^f$ for some $f \in A(K_\varphi)^{**}$ and suppose that $\alpha : \mathcal{N} \to \mathcal{M}$ is a sufficient channel for the set $\{\psi, \varphi\}$. Let us denote $\varphi_0 := \varphi \circ \alpha$, $\psi_0 := \psi \circ \alpha$. Let $\beta : \mathcal{M} \to \mathcal{N}$ be the channel such that $\varphi_0 \circ \beta = \varphi$, $\psi_0 \circ \beta = \psi$. We will show that $\psi_0 = \varphi_0^{\beta(f_\psi)}$. To see this, note that for $\omega_0 \in \mathcal{S}_{\varphi_0}$,

$$\beta(f_\psi)(\omega_0) = f_\psi(\omega_0 \circ \beta) = S(\omega_0 \circ \beta, \psi) - S(\omega_0 \circ \beta, \varphi) - S(\psi, \varphi).$$

Then

$$\begin{aligned} &\beta(f_\psi)(\omega_0) + S(\omega_0, \varphi_0) \\ &= S(\omega_0 \circ \beta, \psi) + S(\omega_0, \varphi_0) - S(\omega_0 \circ \beta, \varphi_0 \circ \beta) - S(\psi, \varphi) \geq -S(\psi, \varphi) = c_\varphi(f_\psi) \end{aligned}$$

by positivity and monotonicity of the relative entropy, and

$$\beta(f_\psi)(\psi_0) + S(\psi_0, \varphi_0) = -S(\psi, \varphi)$$

so that $c_{\varphi_0}(\beta(f_\psi)) = c_\varphi(f_\psi)$ and $\psi_0 = \varphi_0^{\beta(f_\psi)}$.

On the other hand, this implies by Theorem 16.2 that $f_{\psi_0} \in A(K_{\varphi_0})^{**}$ and we obtain in the same way that $\psi = \varphi^{\alpha(f_{\psi_0})}$ and $c_\varphi(\alpha(f_{\psi_0})) = c_{\varphi_0}(f_{\psi_0})$.

**Theorem 16.9** *Let $\psi$ be such that $\partial_\varphi(\psi) \neq \emptyset$ and let $\alpha : \mathcal{N} \to \mathcal{M}$ be a channel. Let $\varphi_0 = \varphi \circ \alpha$, $\psi_0 = \psi \circ \alpha$ The following are equivalent*

(i) *$\alpha$ is sufficient for the pair $\{\varphi, \psi\}$,*
(ii) *$f_{\psi_0} \in A(K_{\varphi_0})^{**}$ and $\psi = \varphi^{\alpha(f_{\psi_0})}$,*
(iii) *$c_{\varphi_0}(f_{\psi_0}) = c_\varphi(f_\psi)$.*

*Proof* The implication (i) $\to$ (ii) was already proved above. Suppose (ii) holds, then

$$\begin{aligned} c_\varphi(\alpha(f_{\psi_0})) &= \alpha(f_{\psi_0})(\psi) + S(\psi, \varphi) = \\ &= -S(\psi_0, \varphi_0) - S(\varphi_0, \psi_0) + S(\psi, \varphi). \end{aligned}$$

By putting $\omega = \varphi$ in Lemma 16.2, we obtain $c_\varphi(\alpha(f_{\psi_0})) \leq -S(\varphi, \psi)$. Then

$$0 \leq S(\psi, \varphi) - S(\psi_0, \varphi_0) \leq S(\varphi_0, \psi_0) - S(\varphi, \psi) \leq 0.$$

It follows that $c_\varphi(f_\psi) = -S(\varphi, \psi) = -S(\varphi_0, \psi_0) = c_{\varphi_0}(f_{\psi_0})$, hence (iii) holds. The implication (iii) $\to$ (i) follows from Theorem 16.7. □

In particular, if $\psi = \varphi^f$ for $f \in A(K_\varphi)$, the above theorem can be formulated as follows.

**Theorem 16.10** *Let $\psi = \varphi^f$, $f \in A(K_\varphi)$ and $\alpha : \mathcal{N} \to \mathcal{M}$ be a channel. Then $\alpha$ is sufficient for $\{\psi, \varphi\}$ if and only if there is some $g_0 \in A(K_{\varphi_0})$, such that $\psi_0 = \varphi_0^{g_0}$ and $f = \alpha(g_0)$.*

*Proof* The statement follows from Theorem 16.9 and the fact that if $\psi = \varphi^f$ for $f \in A_0(K_\varphi)$, then we must have $f = f_\psi$, by Theorem 16.5. □

**Acknowledgement**

Supported by the Center of Excellence EU-QUTE and SAS-Quantum Technologies.

## References

Amari, S. (1985). *Differential-geometrical Methods in Statistics* (New York, Springer-Verlag).

Amari, S. and Nagaoka, H. (2000). *Methods of Information Geometry* (AMS monograph, Oxford University Press).

Asimow, L. and Ellis, A. J. (1980). *Convexity Theory and its Applications in Functional Analysis* (London, Academic Press).

Grasselli, M. R. (2009). Dual connections in nonparametric classical information geometry, *Annals of the Institute of Statistical Mathematics* (to appear) (available at arXiv:math-ph/0104031v1).

Jenčová, A. (2006). A construction of a nonparametric quantum information manifold, *Journal of Functional Analysis* **239**, 1–20.

Jenčová, A. and Petz, D. (2006a). Sufficiency in quantum statistical inference, *Communications in Mathematical Physics* **263**, 259–76.

Jenčová, A. and Petz, D. (2006b). Sufficiency in quantum statistical inference. A survey with examples, *Infinite Dimensional Analysis, Quantum Probability and Related Topics* **9**, 331–51.

Ohya, M. and Petz, D. (1993). *Quantum Entropy and Its Use* (Heidelberg, Springer-Verlag).

Petz, D. (1986). Sufficient subalgebras and the relative entropy of states of a von Neumann algebra, *Communications in Mathematical Physics* **105**, 123–31.

Pistone, G. and Rogantin, M. P. (1999). The exponential statistical manifold, mean parameters, orthogonality and space transformations, *Bernoulli* **5**, 721–60.

Pistone, G. and Sempi, C. (1995). An infinite-dimensional geometric structure on the space of all the probability measures equivalent to the given one, *Annals of Statistics* **23**, 1543–61.

Streater, R. F. (2004). Quantum Orlicz spaces in information geometry, *Open Systems and Information Dynamics* **11**, 359–75.

# Chapter 6

# Comparison of channels and statistical experiments


[CE1] A. Jenčová, Quantum hypothesis testing and sufficient subalgebras, *Lett. Math. Phys.* **93** (2010), 15

[CE2] A. Jenčová, Reversibility conditions for quantum operations, *Rev. Math. Phys.* **24** (2012), 1250016

[CE3] A. Jenčová, Comparison of quantum binary experiments, *Rep. Math. Phys.* **70** (2012), 237-249

[CE4] A. Jenčová, Comparison of quantum channels and quantum statistical experiments, 2016 IEEE International Symposium on Information Theory (ISIT), 2249 - 2253, *IEEE Conference Publications*, 2016

Lett Math Phys (2010) 93:15–27
DOI 10.1007/s11005-010-0398-0

# Quantum Hypothesis Testing and Sufficient Subalgebras

ANNA JENČOVÁ
*Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, 814 73 Bratislava, Slovakia. e-mail: jenca@mat.savba.sk*

Received: 28 October 2008 / Revised: 1 December 2009 / Accepted: 16 May 2010
Published online: 4 June 2010 – © Springer 2010

**Abstract.** We introduce a new notion of a sufficient subalgebra for quantum states: a subalgebra is 2-sufficient for a pair of states $\{\rho_0, \rho_1\}$ if it contains all Bayes optimal tests of $\rho_0$ against $\rho_1$. In classical statistics, this corresponds to the usual definition of sufficiency. We show this correspondence in the quantum setting for some special cases. Furthermore, we show that sufficiency is equivalent to 2-sufficiency, if the latter is required for $\{\rho_0^{\otimes n}, \rho_1^{\otimes n}\}$, for all $n$.

**Mathematics Subject Classification (2000).** 46L53, 81R15, 62B05.

**Keywords.** quantum hypothesis testing, sufficient subalgebras, 2-sufficiency, quantum Chernoff bound.

## 1. Introduction

In order to motivate our results, let us consider the following problem of classical statistics. Suppose that $P_0$ and $P_1$ are two probability distributions and the task is to discriminate between them by an $n$-dimensional observation vector $X$. The problem is, if there is a function (statistic) $T : X \to Y$, such that the vector $Y = T(X)$ (usually of lower dimension) contains all information needed for the discrimination.

In the setting of hypothesis testing, the null hypothesis $H_0 = P_0$ is tested against the alternative $H_1 = P_1$. In the most general formulation, a test is a measurable function $\varphi : X \to [0, 1]$, which can be interpreted as the probability of rejecting the hypothesis if $x \in X$ occurs. There are two kinds of errors appearing in hypothesis testing: it may happen that $H_0$ is rejected, although it is true (error of the first kind), or that it is not rejected when $H_1$ is true (error of the second kind).

Supported by the Slovak Research and Development Agency under the contract No. APVV 0071-06, grant VEGA 2/0032/09, Center of Excellence SAS-Quantum Technologies and ERDF OP R&D Project CE QUTE ITMS 26240120009.

For a given test $\varphi$, the error probabilities are

$$\alpha(\varphi) = \int \varphi(x) P_0(\mathrm{d}x) \quad \text{first kind}$$
$$\beta(\varphi) = \int (1-\varphi(x)) P_1(\mathrm{d}x) \quad \text{second kind}$$

The two kinds of errors are in some sense complementary and it is usually not possible to minimize both error probabilities simultaneously. In the Bayesian approach, we choose a prior probability distribution $\{\lambda, 1-\lambda\}$, $\lambda \in [0, 1]$ on the two hypotheses and then minimize the average (Bayes) error probability

$$\int \varphi(x)\lambda P_0(\mathrm{d}x) + \int (1-\varphi(x))(1-\lambda) P_1(\mathrm{d}x) = \lambda\alpha(\varphi) + (1-\lambda)\beta(\varphi).$$

Suppose now that $T$ is a sufficient statistic for $\{P_0, P_1\}$. Roughly speaking, this means that there exists a common version of the conditional expectation $E[\cdot|T] = E_{P_0}[\cdot|T]$, $P_0$- a.s. and $E[\cdot|T] = E_{P_1}[\cdot|T]$, $P_1$- a.s. If $\varphi$ is any test, then $E[\varphi|T]$ is another test having the same error probabilities. It follows that we can always have an optimal test that is a function of $T$, so that only values of $T(X)$ are needed for optimal discrimination between $P_0$ and $P_1$.

The following theorem states that this can happen if and only if $T$ is sufficient, so that the above property characterizes sufficient statistics. The theorem was proved by Pfanzagl, see also [16].

THEOREM 1. [15] *Let $T : X \to Y$ be a statistic. The following are equivalent.*

1. *For any $\lambda \in (0, 1)$ and any test $\varphi : X \to [0, 1]$, there exists a test $\psi : Y \to [0, 1]$, such that*

$$\lambda\alpha(\psi \circ T) + (1-\lambda)\beta(\psi \circ T) \leq \lambda\alpha(\varphi) + (1-\lambda)\beta(\varphi)$$

2. *$T$ is a sufficient statistic for $\{P_0, P_1\}$.*

The problem of hypothesis testing can be considered also in the quantum setting. Here we deal with a pair of density operators $\rho_0, \rho_1 \in B(\mathcal{H})$, where $\mathcal{H}$ is a finite dimensional Hilbert space and all tests are given by operators $0 \leq M \leq 1$, $M \in B(\mathcal{H})$. The problem of finding the optimal tests (the quantum Neyman–Pearson tests) and average error probabilities was solved by Helstrom and Holevo [6,8].

Here a question arises, if it is possible to discriminate the states optimally by measuring on a given subsystem. Then we can gain some information only on the restricted densities, which, in general, can be distinguished with less precision.

Let $M_0 \subseteq B(\mathcal{H})$ be the subalgebra describing the subsystem we have access to. The average error probabilities for tests in $M_0$ are usually higher than the optimal ones. We will consider the situation that this does not happen and $M_0$ contains some optimal tests for all prior probabilities. In agreement with classical terminology (see [16]), such a subalgebra will be called sufficient with respect to testing problems, or 2-sufficient, for $\{\rho_0, \rho_1\}$.

The quantum counterpart of sufficiency was introduced and studied by Petz, see Chap. 9. in [13], in a more general context. According to this definition, the subalgebra $M_0$ is sufficient for $\{\rho_0, \rho_1\}$, if there exists a completely positive, trace preserving map $M_0 \to B(\mathcal{H})$, that maps both restricted densities to the original ones. Then the restriction to $M_0$ preserves all information needed for discrimination between the states and it is quite easy to see that a sufficient subalgebra must be 2-sufficient.

The conditions for sufficiency seem to be quite restrictive (see for example the factorization conditions in [9]) and might be too strong, if only hypothesis testing is considered. It is therefore natural to ask if there is a quantum version of Theorem 1, that is, if every 2-sufficient subalgebra must be sufficient.

In this paper, we give a partial answer to this question. We show that 2-sufficiency and sufficiency are equivalent under each of the following conditions: (1) the subalgebra $M_0$ is invariant under the modular group of one of the states, (2) $M_0$ is commutative, (3) $\rho_0$ and $\rho_1$ commute. Moreover, we show that if the 2-sufficiency condition is strengthened to hold for $n$ independent copies of the densities for all $n$, then the two notions become equivalent.

The organization of the paper is as follows. In Section 2, some basic notions are introduced and several characterizations of a sufficient subalgebra are given. A new characterization, based on a version of the Radon–Nikodym derivative, is found, this will be needed for the main results. Section 3 gives the quantum Neyman–Pearson lemma and quantum Chernoff bound. Section 4 contains the main results: a convenient necessary condition for 2-sufficiency is found and it is shown that it implies sufficiency in the three above described cases. Finally, the quantum Chernoff bound is utilized to treat the case when 2-sufficiency holds for $n$ independent copies of the states, for all $n$.

## 2. Some Basic Definitions and Facts

### 2.1. GENERALIZED CONDITIONAL EXPECTATION

Let $\mathcal{H}$ be a finite-dimensional Hilbert space and let $\rho$ be an invertible density matrix. Let $M_0 \subseteq B(\mathcal{H})$ be a subalgebra and let $E: B(\mathcal{H}) \to M_0$ be the trace preserving conditional expectation. Then $E(\rho)$ is the restricted density of the state $\rho$.

As we have seen, the classical sufficient statistic is defined by certain property of the conditional expectations. It is well known that in the quantum case, a state preserving conditional expectation does not always exist. Therefore we need the generalized conditional expectation, defined by Accardi and Cecchini [1]. In our setting, it can be given as follows.

Let us introduce the inner product $\langle X, Y\rangle_\rho = \mathrm{Tr}\, X^*\rho^{1/2}Y\rho^{1/2}$ in $B(\mathcal{H})$. Then the generalized conditional expectation $E_\rho$ is a map $B(\mathcal{H}) \to M_0$, defined by

$$\langle X_0, Y\rangle_\rho = \langle X_0, E_\rho(Y)\rangle_{E(\rho)}, \quad X_0 \in M_0,\ Y \in B(\mathcal{H})$$

It is easy to see that we have

$$E_\rho(X) = E(\rho)^{-1/2} E\left(\rho^{1/2} X \rho^{1/2}\right) E(\rho)^{-1/2} \tag{1}$$

It is known that $E_\rho$ is completely positive and unital and that it is a conditional expectation if and only if $\rho^{it} M_0 \rho^{-it} \subseteq M_0$, for all $t \in \mathbb{R}$. It is also easy to see that $E_\rho$ preserves the state $\rho$, that is, $E_\rho^* \circ E(\rho) = \rho$.

Next we introduce two subalgebras, related to $E_\rho$. Let $F_\rho$ be the set of fixed points of $E_\rho$ and let $N_\rho \subseteq B(\mathcal{H})$ be the multiplicative domain of $E_\rho$,

$$N_\rho = \left\{X \in B(\mathcal{H}), E_\rho(X^*X) = E_\rho(X)^* E_\rho(X), E_\rho(XX^*) = E_\rho(X) E_\rho(X)^*\right\}$$

Then both $F_\rho$ and $N_\rho$ are subalgebras in $B(\mathcal{H})$. It is clear that $F_\rho \subseteq M_0 \cap N_\rho$, moreover, $X \in F_\rho$ if and only if $\rho^{it} X \rho^{-it} \in M_0$ for all $t \in \mathbb{R}$. As for $N_\rho$, we have the following result.

LEMMA 1. $N_\rho = \rho^{1/2} M_0 \rho^{-1/2} \cap \rho^{-1/2} M_0 \rho^{1/2}$

*Proof.* It is clear from (1) that $X \in N_\rho$ if and only if

$$E\left(\rho^{1/2} X^* X \rho^{1/2}\right) = E\left(\rho^{1/2} X^* \rho^{1/2}\right) E(\rho)^{-1} E\left(\rho^{1/2} X \rho^{1/2}\right)$$
$$E\left(\rho^{1/2} X X^* \rho^{1/2}\right) = E\left(\rho^{1/2} X \rho^{1/2}\right) E(\rho)^{-1} E\left(\rho^{1/2} X^* \rho^{1/2}\right)$$

Let $A = X\rho^{1/2}$, $B = \rho^{1/2}$. Similarly as in [11], we put $M = A - B\Lambda$, with $\Lambda = E(\rho)^{-1} E(\rho^{1/2} X \rho^{1/2})$. Then from $E(M^*M) \geq 0$, we obtain

$$E(A^*A) \geq E(A^*B) E(\rho)^{-1} E(B^*A),$$

with equality if and only if $M = 0$, this implies

$$\rho^{-1/2} X \rho^{1/2} = E(\rho)^{-1} E\left(\rho^{1/2} X \rho^{1/2}\right) \in M_0.$$

Conversely, let $X_0 = \rho^{-1/2} X \rho^{1/2} \in M_0$, then $E(\rho^{1/2} X \rho^{1/2}) = E(\rho) X_0$, this implies that $M = 0$.

Similarly, we get that $\rho^{-1/2} X^* \rho^{1/2} \in M_0$ is equivalent with the second equality. □

It is also known that $E_\rho(XY) = E_\rho(X) E_\rho(Y)$, $E_\rho(YX) = E_\rho(Y) E_\rho(X)$ for all $X \in N_\rho$, $Y \in B(\mathcal{H})$, this can be also shown from the above Lemma. Note that in the case that $E_\rho$ is a conditional expectation, $F_\rho = N_\rho = M_0$.

2.2. A RADON–NIKODYM DERIVATIVE AND RELATIVE ENTROPIES

Let $\rho_0, \rho_1$ be invertible density matrices in $B(\mathcal{H})$. We will use the quantum version of the Radon–Nikodym derivative introduced in [5]. In our setting, the derivative $d_{\rho_0,\rho_1}$ of $\rho_1$ with respect to $\rho_0$ is defined as the unique element in $B(\mathcal{H})$, such that $\mathrm{Tr}\,\rho_1 X = \langle X^*, d_{\rho_0,\rho_1}\rangle_{\rho_0}$. Then clearly

$$d_{\rho_0,\rho_1} = \rho_0^{-1/2}\rho_1\rho_0^{-1/2}$$

so that $d_{\rho_0,\rho_1}$ is positive, and $\|d_{\rho_0,\rho_1}\| \leq \lambda$ for any $\lambda > 0$, such that $\rho_1 \leq \lambda\rho_0$. It is also easy to see that

$$E_{\rho_0}(d_{\rho_0,\rho_1}) = d_{E(\rho_0),E(\rho_1)}$$

Let us recall that the Belavkin–Staszewski relative entropy is defined as [5]

$$S_{BS}(\rho_1,\rho_0) = -\mathrm{Tr}\,\rho_0\eta\left(\rho_0^{-1/2}\rho_1\rho_0^{-1/2}\right) = -\mathrm{Tr}\,\rho_0\eta(d_{\rho_0,\rho_1})$$

where $\eta(x) = -x\log(x)$. Let $S$ be the Umegaki relative entropy

$$S(\rho_1,\rho_0) = \mathrm{Tr}\,\rho_1(\log\rho_1 - \log\rho_0)$$

then $S(\rho_1,\rho_0) \leq S_{BS}(\rho_1,\rho_0)$, [7] and $S(\rho_1,\rho_0) = S_{BS}(\rho_1,\rho_0)$ if $\rho_0$ and $\rho_1$ commute. Both relative entropies are monotone in the sense that

$$S(\rho_1,\rho_0) \geq S(E(\rho_1),E(\rho_0)), \quad S_{BS}(\rho_1,\rho_0) \geq S_{BS}(E(\rho_1),E(\rho_0))$$

holds for any subalgebra $M_0$. As we will see in the next section, equality in the monotonicity for $S$ is equivalent with sufficiency of the subalgebra $M_0$ with respect to $\{\rho_0,\rho_1\}$. For $S_{SB}$, we have the following result.

LEMMA 2. *The following are equivalent.*

(i) $S_{BS}(\rho_1,\rho_0) = S_{BS}(E(\rho_1),E(\rho_0))$
(ii) $d_{\rho_0,\rho_1} \in N_{\rho_0}$
(iii) $\rho_1\rho_0^{-1} \in M_0$
(iv) $\rho_1\rho_0^{-1} = E(\rho_1)E(\rho_0)^{-1}$

*Proof.* Since the function $-\eta(x) = x\log(x)$ is operator convex,

$$\eta\left(d_{E(\rho_0),E(\rho_1)}\right) = \eta\left(E_{\rho_0}(d_{\rho_0,\rho_1})\right) \leq E_{\rho_0}\left(\eta(d_{\rho_0,\rho_1})\right) \tag{2}$$

by Jensen's inequality. We have

$$\mathrm{Tr}\,\rho_0(E_{\rho_0}(\eta(d_{\rho_0,\rho_1})) - \eta(E_{\rho_0}(d_{\rho_0,\rho_1}))) = S_{BS}(\rho_1,\rho_0) - S_{BS}(E(\rho_1),E(\rho_0))$$

and since $\rho_0$ is invertible, equality in the monotonicity of $S_{BS}$ is equivalent with equality in (2). As it was proved in [14], this happens if and only if $d_{\rho_0,\rho_1} \in N_{\rho_0}$. This shows the equivalence (i) $\leftrightarrow$ (ii). The equivalence of (ii) and (iii) follows by Lemma 1, (iii) $\iff$ (iv) is rather obvious. $\square$

### 2.3. SUFFICIENT SUBALGEBRAS

We say that the subalgebra $M_0 \subseteq B(\mathcal{H})$ is sufficient for $\{\rho_0, \rho_1\}$ if there is a completely positive trace preserving map $T: M_0 \to B(\mathcal{H})$, such that $T \circ E(\rho_0) = \rho_0$ and $T \circ E(\rho_1) = \rho_1$. The following characterizations of sufficiency were obtained by Petz.

THEOREM 2. [10,13] *The following are equivalent.*

(i) $M_0 \subseteq B(\mathcal{H})$ *is sufficient for* $\{\rho_0, \rho_1\}$
(ii) $S(\rho_1, \rho_0) = S(E(\rho_1), E(\rho_0))$
(iii) $\mathrm{Tr}\, \rho_0^s \rho_1^{1-s} = \mathrm{Tr}\, E(\rho_0)^s E(\rho_1)^{1-s}$ *for some* $s \in (0, 1)$
(iv) $\mathrm{Tr}\, E_{\rho_0}(X)\rho_1 = \mathrm{Tr}\, X\rho_1$ *for all* $X \in B(\mathcal{H})$
(v) $E_{\rho_0} = E_{\rho_1}$.

The next characterization is based on the Radon–Nikodym derivative.

THEOREM 3. *The subalgebra* $M_0 \subseteq B(\mathcal{H})$ *is sufficient for* $\{\rho_0, \rho_1\}$ *if and only if* $d_{\rho_0,\rho_1} \in F_{\rho_0}$.

*Proof.* Let us denote $d = d_{\rho_0,\rho_1}$ and $d_0 = d_{E(\rho_0),E(\rho_1)}$. Since $d_0 \in M_0$, we have by definition that

$$\mathrm{Tr}\, \rho_1 E_{\rho_0}(X) = \langle d_0, E_{\rho_0}(X)\rangle_{E(\rho_0)} = \langle d_0, X\rangle_{\rho_0}$$

so that $\mathrm{Tr}\, \rho_1 E_{\rho_0}(X) = \mathrm{Tr}\, \rho_1 X$ if and only if $\langle d_0, X\rangle_{\rho_0} = \langle d, X\rangle_{\rho_0}$. It follows that $d = d_0$ is equivalent with sufficiency of $M_0$, by Theorem 2 (iv). Since $E_{\rho_0}(d) = d_0$, this is equivalent with $d_{\rho_0,\rho_1} \in F_{\rho_0}$. $\square$

## 3. Quantum Hypothesis Testing

Let us now turn to the problem of hypothesis testing. Any test of the hypothesis $H_0 = \rho_0$ against the alternative $H_1 = \rho_1$ is represented by an operator $0 \leq M \leq 1$, which corresponds to rejecting the hypothesis. Then we have the error probabilities

$$\alpha(M) = \mathrm{Tr}\, \rho_0 M \quad \text{first kind}$$
$$\beta(M) = \mathrm{Tr}\, \rho_1(1 - M) \quad \text{second kind}$$

For $\lambda \in (0, 1)$, we define the Bayes optimal test to be a minimizer of the expression

$$\lambda\alpha(M) + (1 - \lambda)\beta(M) \tag{3}$$

It is clear that minimizing (3) is the same as maximizing

$$\mathrm{Tr}\, (\rho_1 - t\rho_0)M, \quad t = \frac{\lambda}{1 - \lambda}$$

### 3.1. THE QUANTUM NEYMAN–PEARSON LEMMA

The following is the quantum version of the Neyman–Pearson lemma. The obtained optimal tests are called the (quantum) Neyman–Pearson tests. We give a simple proof for completeness.

LEMMA 3. *Let $t \geq 0$ and let us denote $P_{t,+} := \mathrm{supp}\,(\rho_1 - t\rho_0)_+$, $P_{t,-} := \mathrm{supp}\,(\rho_1 - t\rho_0)_-$ and $P_{t,0} := 1 - P_{t,+} - P_{t,-}$. Then the operator $0 \leq M_t \leq 1$ is a Bayes optimal test of $\rho_0$ against $\rho_1$ if and only if*

$$M_t = P_{t,+} + X_t$$

*where $0 \leq X_t \leq P_{t,0}$.*

*Proof.* Let $0 \leq M \leq 1$, then

$$\begin{aligned}\mathrm{Tr}\,(\rho_1 - t\rho_0)M &= \mathrm{Tr}\,(\rho_1 - t\rho_0)_+ M - \mathrm{Tr}\,(\rho_1 - t\rho_0)_- M \leq \mathrm{Tr}\,(\rho_1 - t\rho_0)_+ M \\ &\leq \mathrm{Tr}\,(\rho_1 - t\rho_0)_+ = \mathrm{Tr}\,(\rho_1 - t\rho_0) P_{t,+} \qquad (4)\end{aligned}$$

It follows that $M_t = P_{t,+} + X_t$, $X_t \leq P_{t,0}$ is a Bayes optimal test. Conversely, let $M_t$ be some Bayes optimal test, then we must have

$$\mathrm{Tr}\,(\rho_1 - t\rho_0)M_t = \mathrm{Tr}\,(\rho_1 - t\rho_0)_+ M_t = \mathrm{Tr}\,(\rho_1 - t\rho_0) P_{t,+}$$

so that $\mathrm{Tr}\,(\rho_1 - t\rho_0)_- M_t = 0$. By positivity, this implies that $P_{t,-}M_t = M_t P_{t,-} = 0$, so that

$$M_t(P_{t,+} + P_{t,0}) = (P_{t,+} + P_{t,0})M_t = M_t$$

which is equivalent with $M_t \leq P_{t,+} + P_{t,0}$. Furthermore, from

$$\mathrm{Tr}\,(\rho_1 - t\rho_0)_+(P_{t,+} + P_{t,0} - M_t) = 0$$

we obtain $P_{t,+} - P_{t,+}M_t P_{t,+} = P_{t,+}(1 - M_t)P_{t,+} = 0$, hence $(1 - M_t)P_{t,+} = 0$. We obtain $P_{t,+} \leq M_t$ and by putting $X_t := M_t - P_{t,+}$, we get the result. □

Let us denote by $\Pi_{e,\lambda}$ the minimum Bayes error probability. Then

$$\begin{aligned}\Pi_{e,\lambda} &= \lambda\alpha(M_{\lambda/(1-\lambda)}) + (1-\lambda)\beta(M_{\lambda/(1-\lambda)}) \\ &= \frac{1}{2}(1 - \|(1-\lambda)\rho_1 - \lambda\rho_0\|_1) \qquad (5)\end{aligned}$$

where the last equality follows from

$$1 - t = \mathrm{Tr}\,(\rho_1 - t\rho_0) = \mathrm{Tr}\,(\rho_1 - t\rho_0)_+ - \mathrm{Tr}\,(\rho_2 - t\rho_0)_-$$

and

$$\|\rho_1 - t\rho_0\|_1 = \mathrm{Tr}\,|\rho_1 - t\rho_0| = \mathrm{Tr}\,(\rho_1 - t\rho_0)_+ + \mathrm{Tr}\,(\rho_2 - t\rho_0)_-$$

### 3.2. THE QUANTUM CHERNOFF BOUND

Suppose now that we have $n$ copies of the states $\rho_0$ and $\rho_1$, so that we test the hypothesis $\rho_0^{\otimes n}$ against $\rho_1^{\otimes n}$ by means of an operator $0 \leq M_n \leq 1$, $M_n \in \mathcal{B}(\mathcal{H}^{\otimes n})$. Again, we may use the Neyman–Pearson lemma to find the minimum Bayes error probability

$$\Pi_{e,\lambda,n} = \frac{1}{2}\left(1 - \|(1-\lambda)\rho_1^{\otimes n} - \lambda\rho_0^{\otimes n}\|_1\right)$$

The following important result, obtained in [3] and [12] (see also [4]), is the quantum version of the classical Chernoff bound:

$$\lim_n \left(-\frac{1}{n}\log \Pi_{e,\lambda,n}\right) = -\log\left(\inf_{0\leq s\leq 1} \operatorname{Tr}\rho_0^{1-s}\rho_1^s\right) =: \xi_{QCB}(\rho_0, \rho_1) \tag{6}$$

The expression $\xi_{QCB}$ has a number of interesting properties. For example, it was proved that it is always nonnegative and equal to 0 if and only if $\rho_0 = \rho_1$, moreover, it is monotone in the sense that

$$\xi_{QCB}(\rho_0, \rho_1) \geq \xi_{QCB}(E(\rho_0), E(\rho_1))$$

Therefore, although it is not symmetric, $\xi_{QCB}$ provides a reasonable distance measure on density matrices, called the quantum Chernoff distance. Note also that in the case that the matrices are invertible, the infimum is always attained in some $s^* \in [0, 1]$.

## 4. 2-Sufficiency

We say that $M_0$ is sufficient with respect to testing problems, or 2-sufficient, for $\{\rho_0, \rho_1\}$ if for any test $M$ and any $\lambda \in (0, 1)$, there is some test $N_\lambda \in M_0$, such that

$$\lambda\alpha(N_\lambda) + (1-\lambda)\beta(N_\lambda) \leq \lambda\alpha(M) + (1-\lambda)\beta(M)$$

It is quite clear that $M_0$ is 2-sufficient if and only if for all $t \geq 0$, we can find a Neyman–Pearson test $M_t \in M_0$. Moreover, suppose that $M_0$ is a sufficient subalgebra for $\{\rho_0, \rho_1\}$ and let $T = E_{\rho_0} = E_{\rho_1}$. Then, if $M_t$ is a Neyman–Pearson test, then $T(M_t) \in M_0$ is a Neyman–Pearson test as well. Hence, a sufficient subalgebra is always 2-sufficient. In this section, we find the opposite implication in some special cases.

LEMMA 4. *$P_{t,0} \neq 0$ if and only if $t$ is an eigenvalue of $d := d_{\rho_0,\rho_1}$. Moreover, the rank of $P_{t,0}$ is equal to multiplicity of $t$.*

*Proof.* By definition,

$$(\rho_1 - t\rho_0)P_{t,0} = \rho_0^{1/2}(d-t)\rho_0^{1/2}P_{t,0} = 0$$

so that $(d-t)\rho_0^{1/2}P_{t,0}\rho_0^{1/2}=0$. Suppose $P_{t,0}\neq 0$, then $t$ is an eigenvalue of $d$ and any vector in the range of $\rho_0^{1/2}P_{t,0}\rho_0^{1/2}$ is an eigenvector. This implies that $r(P_{t,0})=r(\rho^{1/2}P_{t,0}\rho^{1/2})\leq r(F)$, where $F$ is the eigenprojection of $t$.

Conversely, let $t$ be an eigenvalue of $d$ with the eigenprojection $F$, then

$$(\rho_1-t\rho_0)\rho_0^{-1/2}F\rho_0^{-1/2}=\rho_0^{1/2}(d-t)F\rho_0^{-1/2}=0,$$

so that the range of $\rho^{-1/2}F\rho^{-1/2}$ is in the kernel of $\rho_1-t\rho_0$, this implies $r(F)\leq r(P_{t,0})$. □

Let us denote $Q_{t,+}=\operatorname{supp}(E(\rho_1)-tE(\rho_0))_+$, $Q_{t,0}=\ker(E(\rho_1)-tE(\rho_0))$ and let $\Pi^0_{e,\lambda}$ be the minimal Bayes error probability for the restricted densities

$$\Pi^0_{e,\lambda}:=\inf_{M\in M_0}\lambda\alpha(M)+(1-\lambda)\beta(M)=\frac{1}{2}(1-\|(1-\lambda)E(\rho_1)-\lambda E(\rho_0)\|_1)$$

LEMMA 5. *The following are equivalent.*

(i) *The subalgebra $M_0$ is 2-sufficient for $\{\rho_0,\rho_1\}$.*
(ii) $\Pi^0_{e,\lambda}=\Pi_{e,\lambda}$ *for all* $\lambda\in(0,1)$.
(iii) $Q_{t,0}=P_{t,0}$ *and* $Q_{t,+}=P_{t,+}$ *for all* $t\geq 0$.

*Proof.* It is obvious that (i) implies (ii). Suppose (ii) and let us denote $f(t):=\max_{0\leq M\leq 1}\operatorname{Tr}(\rho_1-t\rho_0)M$. If $N_t$ is any Neyman–Pearson test for $\{E(\rho_0),E(\rho_1)\}$, then

$$\operatorname{Tr}(\rho_1-t\rho_0)N_t=\operatorname{Tr}(E(\rho_1)-tE(\rho_0))N_t=f(t),$$

so that $N_t$ is a Neyman–Pearson test for $\{\rho_0,\rho_1\}$ as well. Putting $N_t=Q_{t,+}$ and $N_t=Q_{t,+}+Q_{t,0}$, we get by Lemma 3 that

$$Q_{t,+}=P_{t,+}+X_t,\quad Q_{t,+}+Q_{t,0}=P_{t,+}+Y_t,$$

with $X_t,Y_t\leq P_{t,0}$. This implies that $Q_{t,0}\leq P_{t,0}$ and $Q_{t,+}=P_{t,+}$ if $P_{t,0}=0$.

Let $t$ be an eigenvalue of $d_0$, then $P_{t,0}\geq Q_{t,0}\neq 0$, hence $t$ is also an eigenvalue of $d$, and its multiplicity in $d_0$ is not greater that its multiplicity in $d$. Since the sum of multiplicities must equal to $m=\dim(\mathcal{H})$, we must have $r(Q_{t,0})=r(P_{t,0})$, so that $Q_{t,0}=P_{t,0}$. This implies that $X_t\leq Q_{t,0}$, hence $X_t=0$ and $P_{t,+}=Q_{t,+}$ for all $t$.

The implication (iii) $\to$ (i) is again obvious. □

Note that the condition (ii) is equivalent with

$$\|E(\rho_1)-tE(\rho_0)\|_1\geq\|\rho_1-t\rho_0\|_1,\quad \text{for all } t\geq 0$$

This condition, with $E(\rho_0)$ and $E(\rho_1)$ replaced by arbitrary densities $\sigma_0$ and $\sigma_1$ was studied in [2]. It was shown that for $2\times 2$ matrices, this is equivalent with the existence of a completely positive trace preserving map $T$, such that $T(\rho_0)=\sigma_0$ and $T(\rho_1)=\sigma_1$. In our case, this means that 2-sufficiency implies sufficiency for $2\times 2$ matrices. Since any nontrivial subalgebra in $\mathcal{M}(\mathbb{C}^2)$ is commutative, this agrees with our results below.

The above Lemma gives characterizations of 2-sufficiency, but the conditions are not easy to check. The next Theorem gives a simple necessary condition.

THEOREM 4. *Let $M_0$ be 2-sufficient for $\{\rho_1,\rho_0\}$. Then $d_{\rho_1,\rho_0}\in N_{\rho_0}$.*

*Proof.* By the previous Lemma, we have $P_{t,0}=Q_{t,0}\in M_0$ for all $t$. Let $t_1,\dots,t_k$ be the eigenvalues of $d$ and denote $P_i=P_{t_i,0}$. Then from $(d-t_i)\rho_0^{1/2}P_i=0$ we get

$$d\rho_0^{1/2}\sum_i P_i=\rho_0^{1/2}\sum_i t_iP_i$$

By Lemma 4 and its proof, $\operatorname{supp}(\rho_0^{1/2}P_i\rho_0^{1/2})\le F_i$ and $r(P_i)=r(F_i)$, with $F_i$ the eigenprojection of $t_i$. It follows that $\sum_i\rho_0^{1/2}P_i\rho_0^{1/2}$, and hence also $\sum_i P_i$, is invertible. Therefore,

$$d\rho_0^{1/2}=\rho_0^{1/2}c,\quad c:=\sum_i t_iP_i\left(\sum_j P_j\right)^{-1}$$

that is, $d=\rho_0^{1/2}c\rho_0^{-1/2}$, with $c\in M_0$. Moreover, $d=d^*=\rho_0^{-1/2}c^*\rho_0^{1/2}$, so that $d\in\rho_0^{1/2}M_0\rho_0^{-1/2}\cap\rho_0^{-1/2}M_0\rho_0^{1/2}$. By Lemma 1, this entails that $d\in N_{\rho_0}$. □

THEOREM 5. *Let the subalgebra $M_0$ be 2-sufficient for $\{\rho_0,\rho_1\}$. Then $M_0$ is sufficient for $\{\rho_0,\rho_1\}$ in each of the following cases.*

(1) $\rho_0^{it}M_0\rho_0^{-it}\subseteq M_0$ *for all* $t\in\mathbb{R}$
(2) $M_0$ *is commutative*
(3) $\rho_0$ *and* $\rho_1$ *commute*

*Proof.* (1) By Theorem 4, we have $d\in N_{\rho_0}$. Since $\rho_0^{it}M_0\rho_0^{-it}\subseteq M_0$, we have $d\in N_{\rho_0}=F_{\rho_0}$. By Theorem 3, this implies that $M_0$ is sufficient.

(2) Since $d\in N_{\rho_0}$, we have $S_{BS}(\rho_1,\rho_0)=S_{BS}(E(\rho_1),E(\rho_0))$, by Lemma 2. Since $M_0$ is commutative,

$$S(E(\rho_1),E(\rho_0))=S_{BS}(E(\rho_1),E(\rho_0))=S_{BS}(\rho_1,\rho_0)\ge S(\rho_1,\rho_0)$$

By monotonicity of the relative entropy, this implies $S(\rho_1,\rho_0)=S(E(\rho_1),E(\rho_0))$, so that $M_0$ is sufficient for $\{\rho_0,\rho_1\}$, by Theorem 2 (ii).

(3) Let $M_1$ be the subalgebra generated by all $P_{t,+}$, $t\in\mathbb{R}$. Then $M_1$ is commutative and 2-sufficient for $\{\rho_0,\rho_1\}$, hence sufficient by (2). If $M_0$ is 2-sufficient,

we must have $M_1 \subseteq M_0$ by Lemma 5, so that $M_0$ must be sufficient for $\{\rho_0, \rho_1\}$ as well. □

It is clear from the proof of (1) that 2-sufficiency implies sufficiency whenever $N_{\rho_0} = F_{\rho_0}$ (or, equivalently, $N_{\rho_1} = F_{\rho_1}$). In fact, it can be shown that $N_{\rho_0} = F_{\rho_0}$ whenever $M_0$ is commutative, which gives an alternative proof of (2). Next we give a further example of this situation.

EXAMPLE 1. Let $\mathcal{H} = \mathbb{C}^4$ and let $M_0 = \mathcal{M}(\mathbb{C}^2) \otimes I \subset B(\mathcal{H})$. Let $\rho$ be a block-diagonal density matrix $\rho = \begin{pmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{pmatrix}$, where $\rho_1, \rho_2$ are positive invertible matrices in $\mathcal{M}(\mathbb{C}^2)$, and let $\sigma$ be any density matrix. Suppose that $M_0$ is 2-sufficient for $\{\rho, \sigma\}$.

By Theorem 4, $d_{\sigma,\rho} \in N_\rho$, which by Lemma 2 is equivalent with $\sigma\rho^{-1} \in M_0$. This implies that $\sigma$ must be block-diagonal as well, $\sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$.

By Lemma 5, $P_{t,+} \in M_0$ for all $t \geq 0$, so that $P_{t,+} = \begin{pmatrix} p_t & 0 \\ 0 & p_t \end{pmatrix}$, where $p_t = \operatorname{supp}(\sigma_1 - t\rho_1)_+ = \operatorname{supp}(\sigma_2 - t\rho_2)_+$. Since $p_t$ is a projection in $\mathcal{M}(\mathbb{C}^2)$, we have he following two possibilities: either $p_t = I$ for $t < t_0$ and $p_t = 0$ for $t \geq t_0$, or $p_t$ is one-dimensional for $t$ in some interval $(t_0, t_1)$. Since $\rho = \sigma$ in the first case, we may suppose that the latter is true, so that $p_t$ is a common eigenprojection of $\sigma_1 - t\rho_1$ and $\sigma_2 - t\rho_2$ for $t \in (t_0, t_1)$. It follows that $\sigma_1 - t\rho_1$ commutes with $\sigma_2 - t\rho_2$ for $t \in (t_0, t_1)$, which implies that $\rho_1$ commutes with $\rho_2$.

Let $X \in N_\rho$, then $X = \rho^{1/2} X_0 \rho^{-1/2}$, where both $X_0$, $\rho X_0 \rho^{-1} \in M_0$. Let $X_0 = Y \otimes I \in M_0$, then $\rho X_0 \rho^{-1} \in M_0$ if and only if $\rho_1 Y \rho_1^{-1} = \rho_2 Y \rho_2^{-1}$, that is, $Y$ commutes with $\rho_2^{-1}\rho_1$. If $\rho_2^{-1}\rho_1$ is a constant, then $\rho^{it} M_0 \rho^{-it} \subseteq M_0$, so that $F_\rho = M_0 = N_\rho$. Otherwise, $Y$ must commute with both $\rho_1$ and $\rho_2$ and in this case, $X = \rho^{1/2} X_0 \rho^{-1/2} = X_0 \in F_\rho$.

In conclusion, if $M_0$ is 2-sufficient for $\{\rho, \sigma\}$, we must have $N_\rho = F_\rho$, so that $M_0$ must be a sufficient subalgebra.

Let us now suppose that we have $n$ independent copies of the states, $\rho_0^{\otimes n}$ and $\rho_1^{\otimes n}$. An optimal test for $H_1 : \rho_0^{\otimes n}$ against $H_1 : \rho_1^{\otimes n}$ usually cannot be obtained as the product of optimal tests, but we may ask if there is some optimal test in $M_0^{\otimes n}$. If this is the case for all $\lambda$, we say that $M_0$ is $(2,n)$-sufficient for $\{\rho_0, \rho_1\}$.

THEOREM 6. *The following conditions are equivalent.*

(i) $M_0$ *is* $(2,n)$*-sufficient for* $\{\rho_0, \rho_1\}$*, for all* $n$*.*
(ii) $M_0$ *is a sufficient subalgebra for* $\{\rho_0, \rho_1\}$*.*

*Proof.* Let us denote

$$\Pi^0_{e,\lambda,n} := \frac{1}{2}\left(1 - \|(1-\lambda)E(\rho_1)^{\otimes n} - \lambda E(\rho_0)^{\otimes n}\|_1\right)$$

By Lemma 5 (ii), the condition (i) implies that $\Pi_{e,\lambda,n} = \Pi^0_{e,\lambda,n}$ for all $n$, hence also

$$\lim_n \left(-\frac{1}{n}\log \Pi_{e,\lambda,n}\right) = \lim_n \left(-\frac{1}{n}\log \Pi^0_{e,\lambda,n}\right)$$

By (6), this entails that

$$\inf_{0\le s\le 1} \operatorname{Tr} \rho_0^{1-s}\rho_1^s = \inf_{0\le s\le 1} \operatorname{Tr} E(\rho_0)^{1-s}E(\rho_1)^s$$

By monotonicity, we have $\operatorname{Tr}\rho_0^{1-s}\rho_1^s \le \operatorname{Tr} E(\rho_0)^{1-s}E(\rho_1)^s$ for all $s\in[0,1]$. Suppose that the infimum on the RHS is attained in some $s_0 \in [0,1]$. Then

$$\operatorname{Tr} E(\rho_0)^{1-s_0}E(\rho_1)^{s_0} = \inf_{0\le s\le 1} \operatorname{Tr} \rho_0^{1-s}\rho_1^s \le \operatorname{Tr}\rho_0^{1-s_0}\rho_1^{s_0}.$$

If $s_0 = 0$ or 1, then the quantum Chernoff distance is equal to 0, so that $\rho_0 = \rho_1$ and the subalgebra $M_0$ is trivially sufficient. Otherwise, we must have $\operatorname{Tr} E(\rho_0)^{1-s_0} E(\rho_1)^{s_0} = \operatorname{Tr}\rho_0^{1-s_0}\rho_1^{s_0}$ for $s_0 \in (0,1)$, which implies that $M_0$ is sufficient for $\{\rho_0, \rho_1\}$, by Theorem 2 (iii).

Conversely, let $E_{\rho^{\otimes n}}$ be the generalized conditional expectation $B(\mathcal{H}^{\otimes n}) \to M_0^{\otimes n}$. It is easy to see that for any invertible density matrix $\rho$, $E_{\rho^{\otimes n}} = E_\rho^{\otimes n}$, so that if $E_{\rho_0} = E_{\rho_1}$, then $E_{\rho_0^{\otimes n}} = E_{\rho_1^{\otimes n}}$ for all $n$. Hence if $M_0$ is sufficient for $\{\rho_0,\rho_1\}$, then $M_0^{\otimes n}$ is sufficient for $\{\rho_0^{\otimes n}, \rho_1^{\otimes n}\}$ for all $n$, this implies (i). □

## References


1. Accardi, L., Cecchini, C.: Conditional expectations in von Neumann algebras and a theorem of Takesaki. J. Funct. Anal. **45**, 245–273 (1982)
2. Alberti, P.M., Uhlmann, A.: A problem relating to the positive linear maps on a matrix algebra. Rep. Math. Phys. **18**, 163–176 (1980)
3. Audenaert, K.M.R., Calsamiglia, J., Masanes, L., Munoz-Tapia, R., Acin, A., Bagan, E., Verstraete, F.: Discriminating states: the quantum Chernoff bound. Phys. Rev. Lett. **98**, 160501 (2007)
4. Audenaert, K.M.R., Nussbaum, M., Szkola, A., Verstraete, F.: Asymptotic error rates in quantum hypothesis testing. Commun. Math. Phys. **279**, 251–283 (2008)
5. Belavkin, V.P., Staszewski, P.: $C^*$-algebraic generalizations of relative entropy and entropy. Ann. Ins. Henri Poincaré Sec. A **73**, 51–58 (1982)
6. Helstrom, C.W.: Quantum Detection and Estimation Theory. Academic Press, New York (1976)
7. Hiai, F., Petz, D.: The proper formula for relative entropy and its asymptotics in quantum probability. Commun. Math. Phys. **143**, 99–114 (1991)
8. Holevo, A.S.: On asymptotically optimal hypothesis testing in quantum statistics. Theor. Probab. Appl. **23**, 411–415 (1978)
9. Jenčová, A., Petz, D.: Sufficiency in quantum statistical inference. Commun. Math. Phys. **263**, 259276 (2006)
10. Jenčová, A., Petz, D.: Sufficiency in quantum statistical inference. A survey with examples. IDAQP **9**, 331–351 (2006)
11. Lieb, E.H., Ruskai, M.B.: Some operator inequalities of the Schwarz type. Adv. Math. **12**, 269–273 (1974)

12. Nussbaum, M., Szkola, A.: The Chernoff lower bound for symmetric quantum hypothesis testing. Ann. Stat. **37**(2), 1040–1057 (2009)
13. Ohya, M., Petz, D.: Quantum Entropy and its Use, 2nd edn. Springer, Heidelberg (2004)
14. Petz, D.: On the equality in Jensen's inequality for operator convex functions. Integral Equ. Oper. Theory **9**, 744–747 (1986)
15. Pfanzagl, J.: A characterization of sufficiency by power functions. Metrika **21**, 197–199 (1974)
16. Strasser, H.: Mathematical Theory of Statistics. Statistical Experiments and Asymptotic Decision Theory. Walter de Gruyter, Berlin (1985)

Reviews in Mathematical Physics
Vol. 24, No. 7 (2012) 1250016 (26 pages)

DOI: 10.1142/S0129055X1250016X

# REVERSIBILITY CONDITIONS FOR QUANTUM OPERATIONS

ANNA JENČOVÁ

*Mathematical Institute,*
*Slovak Academy of Sciences,*
*Štefánikova 49, 814 73 Bratislava, Slovakia*
*jenca@mat.savba.sk*

Received 25 July 2011
Revised 26 April 2012
Accepted 22 May 2012
Published 25 June 2012

We give a list of equivalent conditions for reversibility of the adjoint of a unital Schwarz map, with respect to a set of quantum states. A large class of such conditions is given by preservation of distinguishability measures: $F$-divergences, $L_1$-distance, quantum Chernoff and Hoeffding distances. Here we summarize and extend the known results. Moreover, we prove a number of conditions in terms of the properties of a quantum Radon–Nikodym derivative and factorization of states in the given set. Finally, we show that reversibility is equivalent to preservation of a large class of quantum Fisher informations and $\chi^2$-divergences.

*Keywords*: 2-positive maps; Schwarz maps; reversibility; $f$-divergences; Radon–Nikodym derivative; hypothesis testing; quantum Fisher information.

Mathematics Subject Classification 2010: 81P16, 94A17, 62F03

## 1. Introduction

In the mathematical description of quantum mechanics, a quantum mechanical system is represented by a $C^*$-algebra $\mathcal{A} \subseteq B(\mathcal{H})$ of bounded operators on a Hilbert space $\mathcal{H}$. In the case that $\mathcal{H}$ is finite-dimensional, the physical states of the system are represented by density operators, that is, positive operators with unit trace. The evolution of the system is described, in the Schrödinger picture, by a transformation $T$ on the states. Here $T$ is usually required to be a completely positive trace preserving map on the algebra.

Let $\mathcal{S}$ be a set of states, then $\mathcal{S}$ can be seen as carrying some information. If $\mathcal{S}$ undergoes a quantum operation $T$, then some information can be lost. If $\mathcal{S}$ represents a code which is sent through a noisy channel $T : \mathcal{A} \to \mathcal{B}$, then the resulting code $T(\mathcal{S})$ might contain less information than $\mathcal{S}$. In the framework of quantum statistics, $\mathcal{S}$ represents a prior knowledge on the state of the system and the task of the statistician is to make some inference on the true state. But if, say, $\mathcal{S}$ is a family of states on the bipartite system $\mathcal{A} \otimes \mathcal{B}$ and only the system $\mathcal{A}$ is accessible, then the statistician has to work with the restricted states which might be distinguished with less precision. However, it might happen in some situations that the original information can be recovered, in the sense that there is a quantum operation $S$ such that $S \circ T(\sigma) = \sigma$ for all $\sigma \in \mathcal{S}$. In this case we say that $T$ is reversible for $\mathcal{S}$. Such maps are also called sufficient for $\mathcal{S}$, which comes from the well-known notion of sufficiency in classical statistics.

The information loss under quantum operations is expressed in the monotonicity property of distinguishability measures: quantum $f$-divergences [28] like relative entropy, the $L_1$-distance, quantum Chernoff and Hoeffding distances [2], etc., which means that these measures are non-increasing under quantum operations. It is quite clear that if $T$ is reversible for $\mathcal{S}$, then $T$ must preserve all of these measures on $\mathcal{S}$. It was an important observation in [27] that preservation of the relative entropy, along with other equivalent conditions, is equivalent to reversibility. These results were then extended in the papers [29, 15, 16]; see also [24]. The very recent paper [13] extends the monotonicity results to the case that $T$ is the adjoint of a subunital Schwarz map and proves that reversibility is equivalent to preservation of a large class of quantum $f$-divergences, as well as distinguishability measures related to quantum hypothesis testing: the quantum Chernoff and Hoeffding distances. In the present paper, we find conditions for reversibility in terms of the $L_1$-distance and complete the results for the Chernoff and Hoeffding distances and $L_1$-distance for $n$ copies of the states, giving an answer to some of the questions left open in [13]. Moreover, we find a class of quantum Fisher informations, such that preservation of elements in this class is equivalent to reversibility. We also prove reversibility conditions in terms of a quantum Radon–Nikodym derivative, and a quantum version of the factorization theorem of classical statistics.

The various equivalent reversibility conditions are interesting also from the opposite point of view, when we are interested in the equality conditions for the divergences in the first place. This was used, for example, for a characterization of the quantum Markov property [10, 15, 19, 20], conditions for nullity of the quantum discord [8, Lemma 8.12], [6], conditions for strict decrease of Holevo quantity [31] and the equality conditions in certain Minkowski type quantum inequalities and related quantities, [18].

In a preliminary section, we deal with the properties of positive maps, 2-positive maps and Schwarz maps, and their duals with respect to a state. In particular, we find a new characterization of 2-positivity in terms of generalized Schwarz inequality and we show that a unital positive map has the property that its duals with respect to all states are Schwarz maps, if and only if it is 2-positive. Then we proceed to the various reversibility conditions: we list the already known conditions related to $f$-divergences and give an example of a (non-quadratic and strictly convex) operator convex function $f$, such that preservation of the corresponding $f$-divergence does not imply reversibility. Further, we prove reversibility conditions in terms of a quantum Radon–Nikodym derivative and certain factorization conditions on the states. In Secs. 3.4 and 3.5, we deal with the $L_1$-distance, quantum Chernoff and Hoeffding distances. In the last section, we give the reversibility conditions in terms of the quantum Fisher information.

## 2. Preliminaries

Let $\mathcal{H}$ be a finite-dimensional Hilbert space and let $\mathcal{A} \subseteq B(\mathcal{H})$ be a $C^*$-algebra. We denote by $\mathcal{A}^+$ the positive cone in $\mathcal{A}$ and by $\mathcal{S}(\mathcal{A})$ the set of states on $\mathcal{A}$. For $a \in \mathcal{A}^+$, we denote by $\operatorname{supp} a$ the projection onto the support of $a$, that is, $\operatorname{supp} a$ is the smallest projection $p$ satisfying $ap = a$.

A positive linear functional $\tau$ on $\mathcal{A}$ such that $\tau(ab) = \tau(ba)$ for all $a, b \in \mathcal{A}$ (equivalently, $\tau(a^*a) = \tau(aa^*)$ for all $a \in \mathcal{A}$) is called a trace. We will also require that $\tau$ is faithful, then any linear functional $\varphi$ on $\mathcal{A}$ has the form

$$\varphi(a) = \tau(a\rho_\varphi), \quad a \in \mathcal{A}$$

for a unique operator $\rho_\varphi \in \mathcal{A}$, and $\varphi$ is a state if and only if $\rho_\varphi \geq 0$ and $\tau(\rho_\varphi) = 1$. In this case, $\rho_\varphi$ is called the density operator of $\varphi$ with respect to $\tau$. Conversely, any operator $\rho \in \mathcal{A}^+$ with $\tau(\rho) = 1$ defines a state $\varphi_\rho$ on $\mathcal{A}$ with density $\rho$. Moreover, if $\tau$ is faithful, then

$$\langle a, b\rangle_\tau = \tau(a^*b), \quad a, b \in \mathcal{A}$$

defines an inner product in $\mathcal{A}$.

Clearly, $\mathcal{A}$ inherits the trace $\mathrm{Tr} = \mathrm{Tr}_\mathcal{H}$ from $B(\mathcal{H})$, but in general, there exists different faithful traces on $\mathcal{A}$ even if we require $\tau(I) = \mathrm{Tr}(I)$. We will consider general traces only in Sec. 2.4, in the rest of the paper we always assume that $\tau = \mathrm{Tr} = \mathrm{Tr}_\mathcal{H}$ for a fixed representation $\mathcal{A} \subseteq B(\mathcal{H})$. Accordingly, the density operators with respect to Tr will be referred to simply as density operators and we will identify $\mathcal{S}(\mathcal{A})$ with the set $\{\rho \in \mathcal{A}^+, \mathrm{Tr}\,\rho = 1\}$. We will also denote $\langle a, b\rangle := \langle a, b\rangle_{\mathrm{Tr}}$ the restriction of the Hilbert–Schmidt inner product in $B(\mathcal{H})$.

### 2.1. *Positive maps*

Let $\mathcal{B} \subseteq B(\mathcal{K})$ be a finite-dimensional $C^*$ algebra and let $T : \mathcal{A} \to \mathcal{B}$ be a positive map. Let $T^*$ be the adjoint of $T$, with respect to the Hilbert–Schmidt inner product. We will say that $T$ is faithful if $T(a) = 0$ for $a \geq 0$ implies $a = 0$.

**Lemma 1.** *Suppose that $T : \mathcal{A} \to \mathcal{B}$ is a positive map. The following are equivalent.*

(i) *$T(\rho)$ is invertible for any positive invertible $\rho$.*
(ii) *$T(\rho)$ is invertible for some positive invertible $\rho$.*
(iii) *$T^*$ is faithful.*

**Proof.** The implication (i) $\Rightarrow$ (ii) is trivial. Suppose (ii) and let $a \geq 0$ be such that $T^*(a) = 0$. Then $0 = \mathrm{Tr}\, T^*(a)\rho = \mathrm{Tr}\, aT(\rho)$, hence $a = 0$.

Suppose (iii) and let $\rho$ be any positive invertible element. Let $q := \operatorname{supp} T(\rho)$. Then $0 = \operatorname{Tr} T(\rho)(I-q) = \operatorname{Tr} \rho T^*(I-q)$, this implies $I - q = 0$, hence (i) holds. □

**Lemma 2.** *Let $T : \mathcal{A} \to \mathcal{B}$ be a positive map, such that $T^*(I) \leq I$. Let $\rho$ and $\sigma$ be positive operators and let $p = \operatorname{supp} \rho$, $p_0 = \operatorname{supp} T(\rho)$, $q = \operatorname{supp} \sigma$ and $q_0 = \operatorname{supp} T(\sigma)$. Then*

(i) $T^*(I - p_0) \leq I - p$.
(ii) *if* $q \leq p$ *then* $q_0 \leq p_0$.
(iii) $T(p\mathcal{A}p) \subseteq p_0\mathcal{B}p_0$.
(iv) *if* $T^*$ *is unital, then* $T^*(p_0) \geq p$.

**Proof.** Note that for $0 \leq a \leq I$ and any positive $\omega$, $a \leq I - \operatorname{supp} \omega$ if and only if $\operatorname{Tr} a\omega = 0$. We have

$$\operatorname{Tr} \rho T^*(I - p_0) = \operatorname{Tr} T(\rho)(I - p_0) = 0$$

which implies (i). Moreover, suppose $q \leq p$, then by (i),

$$0 \leq \operatorname{Tr} T(\sigma)(I - p_0) = \operatorname{Tr} \sigma T^*(I - p_0) \leq \operatorname{Tr} \sigma(I - p) = 0$$

this proves (ii). Let $a$ be a positive element in $p\mathcal{A}p$, then $\operatorname{supp} a \leq p$, hence by (ii), $\operatorname{supp} T(a) \leq p_0$, so that $T(a) \in p_0\mathcal{B}p_0$. Since $p\mathcal{A}p$ is generated by its positive cone, this implies (iii).

Finally, (iv) follows directly from (i) if $T^*$ is unital. □

We say that $T$ is $n$-positive if the map

$$T_{(n)} := id_n \otimes T : M_n(\mathbb{C}) \otimes \mathcal{A} \to M_n(\mathbb{C}) \otimes \mathcal{B}$$

is positive, and $T$ is completely positive if it is $n$-positive for all $n$. The adjoint $T^*$ is $n$-positive if and only if $T$ is $n$-positive.

### 2.2. *2-positive maps and Schwarz maps*

We say that $T$ is a Schwarz map if it satisfies the Schwarz inequality

$$T(a^*a) \geq T(a)^*T(a), \quad a \in \mathcal{A}. \tag{1}$$

This implies that $T$ is positive and subunital, that is, $T(I) \leq I$. It is well-known that a unital 2-positive map is a Schwarz map [25, Proposition 3.3].

Let $c \in \mathcal{A}^+$ and $a \in \mathcal{A}$. We define $a^*c^{-1}a := \lim_{\varepsilon \to 0} a^*(c + \varepsilon I)^{-1}a$, if the limit exists. Note that this is the case if and only if the range of $a$ is contained in the range of $c$ and then $a^*c^{-1}a = ac^-a$, where $c^-$ denotes the generalized inverse of $c$.

**Lemma 3.** *Let $a, b, c \in \mathcal{A}$. Then the block matrix $M = \begin{pmatrix} a & b \\ b^* & c \end{pmatrix}$ is positive if and only if $c \geq 0$, $bc^{-1}b^*$ is defined and satisfies $a \geq bc^{-1}b^*$.*

**Proof.** The proof for the case that $c$ is invertible can be found in [4]. For the general case, note that $M \geq 0$ if and only if $\left(\begin{smallmatrix} a & b \\ b^* & c+\varepsilon I \end{smallmatrix}\right)$ is positive for all $\varepsilon > 0$. By the first part of the proof, this is equivalent to $c \geq 0$ and $a \geq b(c+\varepsilon I)^{-1}b^*$ for all $\varepsilon > 0$. Since $b(c+\varepsilon I)^{-1}b^*$ is an increasing net of positive operators, the limit exists if and only if it is bounded from above, this proves the lemma. □

Let $c \in \mathcal{A}$ be a positive invertible element. Then we say that $T$ satisfies the generalized Schwarz inequality for $c$ if for all $a \in \mathcal{A}$, $T(a)^*T(c)^{-1}T(a)$ is defined and satisfies [21]

$$T(a^*c^{-1}a) \geq T(a)^*T(c)^{-1}T(a), \quad a \in \mathcal{A}. \tag{2}$$

Note that the condition that $T(a)^*T(c)^{-1}T(a)$ is defined is satisfied if $T^*$ is sub-unital, by Lemma 2(iii).

The next proposition gives a characterization of 2-positivity of maps in terms of the generalized Schwarz inequality, which might be interesting in its own right:

**Proposition 1.** *Let $T : \mathcal{A} \to \mathcal{B}$ be a positive map. Then $T$ is 2-positive if and only if $T$ satisfies the generalized Schwarz inequality for every positive invertible $c \in \mathcal{A}$.*

**Proof.** Let $M = \left(\begin{smallmatrix} a & b \\ b^* & c \end{smallmatrix}\right)$ be a positive element in $M_2(\mathcal{A})$. Let $\varepsilon > 0$ and denote $M_\varepsilon := \left(\begin{smallmatrix} a & b \\ b^* & c+\varepsilon I \end{smallmatrix}\right)$. Then $M_\varepsilon \geq 0$ and it is clear that $T_{(2)}(M) \geq 0$ if and only if $T_{(2)}(M_\varepsilon) \geq 0$ for all $\varepsilon > 0$. Hence we may suppose that $c$ is invertible. In this case, $M \geq 0$ if and only if $c \geq 0$ and $a - bc^{-1}b^* \geq 0$, by Lemma 3. Then

$$M = \begin{pmatrix} a - bc^{-1}b^* & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} bc^{-1}b^* & b \\ b^* & c \end{pmatrix}$$

where both summands are positive. Since $T$ is positive, this implies that $T$ is 2-positive if and only if for all $b \in \mathcal{A}$ and invertible $c \in \mathcal{A}^+$,

$$T_{(2)} \begin{pmatrix} bc^{-1}b^* & b \\ b^* & c \end{pmatrix} = \begin{pmatrix} T(bc^{-1}b^*) & T(b) \\ T(b^*) & T(c) \end{pmatrix} \geq 0.$$

Again by Lemma 3, this is equivalent to the generalized Schwarz inequality for $c$. □

### 2.3. *The map $T_\rho$*

Let $\rho \in \mathcal{S}(\mathcal{A})$. We define a sesquilinear form in $\mathcal{A}$ by

$$\langle a, b \rangle_\rho = \operatorname{Tr} a^* \rho^{1/2} b \rho^{1/2}, \quad a, b \in \mathcal{A}.$$

Then $\langle \cdot, \cdot \rangle_\rho$ defines an inner product in $p\mathcal{A}p$, where $p = \operatorname{supp} \rho$.

Let $T : \mathcal{A} \to \mathcal{B}$ be a positive and trace preserving map, so that $T(\rho)$ is a density operator in $\mathcal{B}$. Let $p_0 = \operatorname{supp} T(\rho)$, then by Lemma 2(iii), $T(p\mathcal{A}p) \subseteq p_0\mathcal{B}p_0$.

The map $T_\rho : p\mathcal{A}p \to p_0\mathcal{B}p_0$ is defined by

$$T_\rho(b) = T(\rho)^{-1/2} T(\rho^{1/2} b \rho^{1/2}) T(\rho)^{-1/2}, \quad b \in p\mathcal{A}p.$$

Note that $T_\rho(a)$ is the unique element in $p_0\mathcal{B}p_0$ satisfying

$$\langle T^*(b), a\rangle_\rho = \langle b, T_\rho(a)\rangle_{T(\rho)}, \quad b \in \mathcal{B}, \tag{3}$$

so that $T_\rho$ is the dual of the unital map $T^*$, defined in [26]. Note also that $T_\rho$ is positive and unital and its adjoint $T_\rho^* : p_0\mathcal{B}p_0 \to p\mathcal{A}p$,

$$T_\rho^*(b) = \rho^{1/2}T^*(T(\rho)^{-1/2}bT(\rho)^{-1/2})\rho^{1/2}$$

satisfies

$$T_\rho^* \circ T(\rho) = \rho \tag{4}$$

by Lemma 2(iv).

It can be shown that $T$ is $n$-positive if and only if $T_\rho$ is $n$-positive. We will now investigate the case when $T_\rho$ is a Schwarz map.

**Lemma 4.** *Let $T : \mathcal{A} \to \mathcal{B}$ be a positive trace preserving map and suppose that $\rho$ is an invertible density operator. Then $T_\rho$ is a Schwarz map if and only if $T$ satisfies the generalized Schwarz inequality for $c = \rho$.*

**Proof.** $T_\rho$ satisfies the Schwarz inequality (1) if and only if

$$T(\rho^{1/2}b^*b\rho^{1/2}) \geq T(\rho^{1/2}b^*\rho^{1/2})T(\rho)^{-1}T(\rho^{1/2}b\rho^{1/2}), \quad b \in \mathcal{A}.$$

Putting $a = \rho^{1/2}b\rho^{1/2}$, we see that this is equivalent to

$$T(a^*\rho^{-1}a) \geq T(a)^*T(\rho)^{-1}T(a), \quad a \in \mathcal{A}. \qquad \square$$

The above lemma, together with Proposition 1, implies the following result. Its importance will become clear at the beginning of Sec. 3.

**Proposition 2.** *Let $T : \mathcal{A} \to \mathcal{B}$ be a positive trace preserving map. Then $T_\rho$ is a Schwarz map for any invertible density operator $\rho$ if and only if $T$ is 2-positive.*

### 2.4. *Multiplicative domain and fixed points*

This section contains some known results on the multiplicative domains and sets of fixed points of unital Schwarz maps and related decompositions of the density operators. We include the proofs partly for the convenience of the reader, and partly because we need a particular form of some of the results (mainly Theorem 2(v) and 2(vi)) which might be difficult to find explicitly in the literature.

Let $\mathcal{B} \subset \mathcal{A} \subseteq B(\mathcal{H})$ be a $C^*$-subalgebra. We will denote by $\mathcal{A}'$ the commutant of $\mathcal{A}$, that is the set of all elements in $B(\mathcal{H})$, commuting with $\mathcal{A}$. Then $\mathcal{A}'$ is a $C^*$-subalgebra in $B(\mathcal{H})$. The relative commutant of $\mathcal{B}$ in $\mathcal{A}$ is the subalgebra $\mathcal{B}' \cap \mathcal{A}$. A conditional expectation $E : \mathcal{A} \to \mathcal{B}$ is a positive linear map, such that $E(bac) = bE(a)c$ for all $a \in \mathcal{A}$, $b, c \in \mathcal{B}$. Such a map is always completely positive. There exists a unique trace preserving conditional expectation $E : \mathcal{A} \to \mathcal{B}$, determined by

$\mathrm{Tr}(ab) = \mathrm{Tr}(E(a)b)$ for $a \in \mathcal{A}$, $b \in \mathcal{B}$ (that is, $E$ is the adjoint of the embedding $\mathcal{B} \hookrightarrow \mathcal{A}$ with respect to $\langle\cdot,\cdot\rangle$).

Let $\Phi : \mathcal{A} \to \mathcal{B}$ be a unital Schwarz map. Let us denote

$$\mathcal{M}_\Phi := \{a \in \mathcal{A},\ \Phi(a^*a) = \Phi(a)^*\Phi(a),\ \Phi(aa^*) = \Phi(a)\Phi(a)^*\}.$$

It is known that [13, Lemma 3.9]

$$\mathcal{M}_\Phi = \{a \in \mathcal{A},\ \Phi(ab) = \Phi(a)\Phi(b),\ \Phi(ba) = \Phi(b)\Phi(a), \forall\, b \in \mathcal{A}\}.$$

This implies that $\mathcal{M}_\Phi$ is a subalgebra in $\mathcal{A}$, called the multiplicative domain of $\Phi$. The restriction of $\Phi$ to $\mathcal{M}_\Phi$ is a *-homomorphism.

Let now $\Phi : \mathcal{A} \to \mathcal{A}$ be a unital Schwarz map and suppose that there is an invertible density operator $\rho \in \mathcal{S}(\mathcal{A})$, such that $\Phi^*(\rho) = \rho$. Let us denote by $\mathcal{F}_\Phi$ the set of fixed points of $\Phi$, that is,

$$\mathcal{F}_\Phi := \{a \in \mathcal{A},\ \Phi(a) = a\}$$

and let $\varphi_\rho$ denote the state $\varphi_\rho(a) = \mathrm{Tr}\,\rho a$ for $a \in \mathcal{A}$.

**Theorem 1.** (i) *$\mathcal{F}_\Phi$ is a subalgebra in $\mathcal{M}_\Phi$.*

(ii) *There exists a conditional expectation $E_\Phi : \mathcal{A} \to \mathcal{F}_\Phi$, such that $E_\Phi^*(\rho) = \rho$.*

(iii) *$\rho^{it}\mathcal{F}_\Phi\rho^{-it} \subseteq \mathcal{F}_\Phi$ for all $t \in \mathbb{R}$.*

(iv) *Let us fix a faithful trace $\tau$ in $\mathcal{F}_\Phi$. Then we have a decomposition*

$$\rho = \rho^A\rho^B,$$

*where $\rho^A \in \mathcal{F}_\Phi$ is the density operator with respect to $\tau$ of the restriction of $\varphi_\rho$ to $\mathcal{F}_\Phi$ and $\rho^B \in \mathcal{F}'_\Phi \cap \mathcal{A}$ is a positive invertible element such that $\Phi^*(\rho^B) = \rho^B$.*

**Proof.** (i) Let $a \in \mathcal{F}_\Phi$, then since $\Phi$ is a Schwarz map, $\Phi(a^*a) \geq \Phi(a)^*\Phi(a) = a^*a$. But we have $\mathrm{Tr}\,\rho(\Phi(a^*a) - a^*a) = 0$, so that $\Phi(a^*a) = a^*a$, similarly $\Phi(aa^*) = aa^*$, hence $a \in \mathcal{M}_\Phi$. Let now $a, b \in \mathcal{F}_\Phi$, then $\Phi(ab) = \Phi(a)\Phi(b) = ab$ and obviously $\Phi(a + b) = a + b$, $\Phi(I) = I$, so that $\mathcal{F}_\Phi$ is a subalgebra.

(ii) Let $E_\Phi := \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1}\Phi^k$, then by the ergodic theorem, $E_\Phi$ is a conditional expectation onto the fixed point subalgebra $\mathcal{F}_\Phi$. It is obvious that $E_\Phi^*(\rho) = \rho$.

(iii) Is equivalent to (ii) by Takesaki's theorem [34].

(iv) It was shown in [15] that for any subalgebra satisfying (iii), there is a decomposition $\rho = \rho^A\rho^B$, where $\rho^A$ is the density of the restriction of $\varphi_\rho$ to $\mathcal{F}_\Phi$ with respect to $\tau$ and $\rho^B$ is a positive invertible element in the relative commutant $\mathcal{F}'_\Phi \cap \mathcal{A}$. For any $a \in \mathcal{A}$,

$$\mathrm{Tr}\,\Phi(a)\rho = \mathrm{Tr}\,\Phi(a)\rho^A\rho^B = \mathrm{Tr}\,\Phi(a\rho^A)\rho^B = \mathrm{Tr}\,a\rho^A\Phi^*(\rho^B)$$

so that $\rho^A\rho^B = \rho = \Phi^*(\rho) = \rho^A\Phi^*(\rho^B)$, this implies $\rho^B = \Phi^*(\rho^B)$. □

**Theorem 2.** *Let $\rho \in \mathcal{A}$ be an invertible density operator and let $T : \mathcal{A} \to \mathcal{B}$ be a trace preserving map, such that both $T^*$ and $T_\rho$ are Schwarz maps and $\rho_0 := T(\rho)$ is invertible. Denote $\Phi := T^* \circ T_\rho$ and $\tilde{\Phi} := T_\rho \circ T^*$. Then*

(i) *$\mathcal{F}_{\tilde{\Phi}}$ is a subalgebra in $\mathcal{M}_{T^*}$ and $\mathcal{F}_\Phi$ is a subalgebra in $\mathcal{M}_{T_\rho}$.*
(ii) *The restriction of $T^*$ is a *-isomorphism from $\mathcal{F}_{\tilde{\Phi}}$ onto $\mathcal{F}_\Phi$, and its inverse is the restriction of $T_\rho$.*
(iii) *$\mathcal{F}_\Phi$ is a subalgebra in $T^*(\mathcal{M}_{T^*})$.*
(iv) *$\rho^{it}\mathcal{F}_\Phi\rho^{-it} \subseteq \mathcal{F}_\Phi$ and $\rho_0^{it}\mathcal{F}_{\tilde{\Phi}}\rho_0^{-it} \subseteq \mathcal{F}_{\tilde{\Phi}}$, for all $t \in \mathbb{R}$.*
(v) *$T(\mathcal{F}'_\Phi \cap \mathcal{A}) \subseteq \mathcal{F}'_{\tilde{\Phi}} \cap \mathcal{B}$*
(vi) *There are decompositions*

$$\rho = T^*(\rho_0^A)\rho^B, \quad \rho_0 = \rho_0^A T(\rho^B)$$

*where $\rho_0^A \in \mathcal{F}_{\tilde{\Phi}}^+$ and $\rho^B \in \mathcal{F}'_\Phi \cap \mathcal{A}^+$ is such that $\Phi^*(\rho^B) = \rho^B$.*

**Proof.** Note that we have $\Phi^*(\rho) = \rho$ and $\tilde{\Phi}^*(\rho_0) = T \circ \Phi^*(\rho) = \rho_0$. Moreover, since $T_\rho^*(\rho_0) = \rho$, $T_\rho$ is faithful by Lemma 1.

By Theorem 1(i), $\mathcal{F}_{\tilde{\Phi}}$ is a subalgebra in $\mathcal{M}_{\tilde{\Phi}}$. It is easy to see that, since $T_\rho$ is faithful, $\mathcal{M}_{\tilde{\Phi}} \subseteq \mathcal{M}_{T^*}$. The second inclusion in (i) is proved similarly.

By (i), the restriction of $T^*$ is a *-homomorphism on $\mathcal{F}_{\tilde{\Phi}}$. Since $\Phi \circ T^* = T^* \circ \tilde{\Phi}$ and $\tilde{\Phi} \circ T_\rho = T_\rho \circ \Phi$, we have $T^*(\mathcal{F}_{\tilde{\Phi}}) \subseteq \mathcal{F}_\Phi$, $T_\rho(\mathcal{F}_\Phi) \subseteq \mathcal{F}_{\tilde{\Phi}}$ and $T_\rho \circ T^*(a) = a$ for $a \in \mathcal{F}_{\tilde{\Phi}}$, this proves (ii).

(iii) Follows from (i) and (ii).

(iv) Follows from Theorem 1(iii).

To prove (v), let $b \in \mathcal{F}_{\tilde{\Phi}}$, $a \in \mathcal{F}'_\Phi \cap \mathcal{A}$ and $c \in \mathcal{B}$. Then

$$\begin{aligned}\operatorname{Tr} cbT(a) &= \operatorname{Tr} T^*(cb)a = \operatorname{Tr} T^*(c)T^*(b)a = \operatorname{Tr} T^*(c)aT^*(b) = \operatorname{Tr} T^*(b)T^*(c)a \\ &= \operatorname{Tr} T^*(bc)a = \operatorname{Tr} bcT(a) = \operatorname{Tr} cT(a)b\end{aligned}$$

so that $T(a) \in \mathcal{F}'_{\tilde{\Phi}} \cap \mathcal{B}$, where we used the fact that $b \in \mathcal{M}_{T^*}$, $T^*(b) \in \mathcal{F}_\Phi$ and cyclicity of the trace.

To prove (vi), let $\tau$ be the restriction of $\operatorname{Tr}$ to $\mathcal{F}_\Phi$. By (ii), $\tilde{\tau} := \tau \circ T^*$ defines a faithful trace on $\mathcal{F}_{\tilde{\Phi}}$. By Theorem 1(iv), we have the decompositions

$$\rho = \rho^A \rho^B, \quad \rho_0 = \rho_0^A \rho_0^B$$

where $\rho^A(\rho_0^A)$ is the density of the restriction of $\varphi_\rho(\varphi_{\rho_0})$ to $\mathcal{F}_\Phi(\mathcal{F}_{\tilde{\Phi}})$ with respect to $\tau$ ($\tilde{\tau}$). Let now $a \in \mathcal{F}_\Phi$, then

$$\begin{aligned}\tau(aT^*(\rho_0^A)) &= \tau(\Phi(a)T^*(\rho_0^A)) = \tau(T^*(T_\rho(a))T^*(\rho_0^A)) = \tau(T^*(T_\rho(a)\rho_0^A)) \\ &= \tilde{\tau}(T_\rho(a)\rho_0^A) = \operatorname{Tr} T_\rho(a)\rho_0 = \operatorname{Tr} a\rho = \tau(a\rho^A).\end{aligned}$$

It follows that $\rho^A = T^*(\rho_0^A)$. If $b \in \mathcal{B}$, then

$$\operatorname{Tr} T^*(b)\rho = \operatorname{Tr} T^*(b)T^*(\rho_0^A)\rho^B = \operatorname{Tr} T^*(b\rho_0^A)\rho^B = \operatorname{Tr} b\rho_0^A T(\rho^B)$$

so that $\rho_0 = \rho_0^A T(\rho^B)$. □

## 3. Conditions for Reversibility

Let $\mathcal{A} \subseteq B(\mathcal{H})$ and $\mathcal{B} \subseteq B(\mathcal{K})$ be finite-dimensional $C^*$-algebras. Let $\mathcal{S} \subset \mathcal{S}(\mathcal{A})$ be a set of density operators and let $T : \mathcal{A} \to \mathcal{B}$ be such that $T^*$ is a unital Schwarz map. We say that $T$ is reversible (or sufficient) for $\mathcal{S}$ if there is a map $S : \mathcal{B} \to \mathcal{A}$, such that $S^*$ is a unital Schwarz map and

$$S \circ T(\sigma) = \sigma, \quad \sigma \in \mathcal{S}. \tag{5}$$

In this section, we study various conditions for reversibility. If not stated otherwise, we assume that the following two conditions hold:

(1) $\mathcal{S}$ contains an invertible element $\rho$ and $T(\rho)$ is invertible as well.
(2) $T : \mathcal{A} \to \mathcal{B}$ is such that both $T^*$ and $T_\rho$ are unital Schwarz maps.

In the original approach of [29], the map $T$ and the recovery map $S$ were both required to be 2-positive. The possibility of weakening this assumption was discussed in [13, Remark 5.8], where the question was raised whether it is enough to assume that $T^*$ is a unital Schwarz map for the map $T_\rho$ to be a Schwarz map as well. Proposition 2 above shows that this is not the case, in fact, it follows that if Condition 2 holds for any density $\rho$, then $T$ must be 2-positive. Moreover, as we will see in Theorem 4, regarding reversibility of $T$, Condition 2 is not more general than assuming that $T$ is a completely positive map.

On the other hand, note that the Condition 1 is not restrictive. Indeed, for $\mathcal{S} \subset \mathcal{S}(\mathcal{A})$ there always exists a (finite) convex combination $\rho$ of elements in $\mathcal{S}$, such that $\operatorname{supp} \sigma \leq \operatorname{supp} \rho =: p$ for all $\sigma \in \mathcal{S}$. Moreover, $T$ is reversible for $\mathcal{S}$ if and only if it is reversible for the closed convex hull $\bar{co}(\mathcal{S})$, therefore, we may always suppose that $\rho \in \mathcal{S}$. By Lemma 2, we also have $p_0 := \operatorname{supp} T(\rho) \geq \operatorname{supp} T(\sigma)$ for all $\sigma \in \mathcal{S}$. Hence $\mathcal{S} \subset \mathcal{S}(p\mathcal{A}p)$ and $T(\mathcal{S}) \subset \mathcal{S}(p_0\mathcal{B}p_0)$.

Let $\tilde{T}$ be the restriction of $T$ to $p\mathcal{A}p$, then $\tilde{T}$ maps $p\mathcal{A}p$ into $p_0\mathcal{B}p_0$, by Lemma 2. We have $\tilde{T}(\sigma) = T(\sigma)$ for $\sigma \in \mathcal{S}$. Again by Lemma 2,

$$\tilde{T}^*(p_0) = pT^*(p_0)p = p,$$

so that $\tilde{T}^*$ is a unital Schwarz map. Note also that $\tilde{T}_\rho = T_\rho$. It follows that if $T$ satisfies Condition 2, then $\tilde{T}$ satisfies both 1 and 2. Moreover, $T$ is reversible for $\mathcal{S} \subset \mathcal{S}(\mathcal{A})$ if and only if $\tilde{T}$ is reversible for $\mathcal{S} \subset \mathcal{S}(p\mathcal{A}p)$. Indeed, let $\tilde{S}$ be the restriction of $S$ to $p_0\mathcal{B}p_0$, where $S : \mathcal{B} \to \mathcal{A}$ is the adjoint of a unital Schwarz map satisfying (5). Then $\tilde{S}$ maps $p_0\mathcal{B}p_0$ into $p\mathcal{A}p$, $\tilde{S}^*$ is a unital Schwarz map and $\tilde{S} \circ \tilde{T}(\sigma) = S \circ T(\sigma) = \sigma$ for all $\sigma \in \mathcal{S}$. Conversely, let $\tilde{S} : p_0\mathcal{B}p_0 \to p\mathcal{A}p$ be the adjoint of a unital Schwarz map, such that $\tilde{S} \circ \tilde{T}(\sigma) = \sigma$ for $\sigma \in \mathcal{S}$, then we extend $\tilde{S}$ to a map $S : \mathcal{B} \to \mathcal{A}$ by

$$S(b) = \tilde{S}(p_0 b p_0) + [\operatorname{Tr} b(1 - p_0)]\rho \quad b \in \mathcal{B}.$$

Then $S^*$ is a unital Schwarz map and $S \circ T(\sigma) = \tilde{S} \circ \tilde{T}(\sigma) = \sigma$ for every $\sigma \in \mathcal{S}$. Moreover, $S$ is $n$-positive whenever $\tilde{S}$ is $n$-positive.

The above constructions can be easily illustrated in the trivial case when $\mathcal{S} = \{\rho\}$. Then both $T$ and $\tilde{T}$ are always reversible, the recovery map being $T^*_\rho$ for $\tilde{T}$, and an extension of $T^*_\rho$ for $T$.

### 3.1. *Quantum f-divergences*

Let $f : [0, \infty) \to \mathbb{R}$ be a function. Recall that $f$ is operator convex if $f(\lambda A + (1 - \lambda)B) \leq \lambda f(A) + (1-\lambda) f(B)$ for any $\lambda \in [0, 1]$ and any positive matrices $A, B$ of any dimension. It was proved in [13] that any operator convex function has an integral representation of the form

$$f(x) = f(0) + ax + bx^2 + \int_{(0,\infty)} \left( \frac{x}{1+t} - \frac{x}{x+t} \right) d\mu_f(t), \quad x \in [0, \infty)$$

where $a \in \mathbb{R}$, $b \geq 0$ and $\mu_f$ is a non-negative measure on $(0, \infty)$ satisfying $\int (1 + t)^{-2} d\mu_f(t) < \infty$.

Let now $\sigma$ and $\rho$ be two density operators and suppose that $\operatorname{supp} \sigma \leq \operatorname{supp} \rho$. Let $\Delta_{\sigma,\rho} = L_\sigma R_\rho^{-1}$ be the relative modular operator, note that $\Delta_{\sigma,\rho}(a) = \sigma a \rho^{-1}$ for any $a \in \mathcal{A}$. Let $f : [0, \infty) \to \mathbb{R}$ be an operator convex function. The $f$-divergence of $\sigma$ with respect to $\rho$ is defined by

$$S_f(\sigma, \rho) = \langle \rho^{1/2}, f(\Delta_{\sigma,\rho}) \rho^{1/2} \rangle$$

see [13] also for the case of arbitrary pairs of density operators. A well-known example is the relative entropy $S(\sigma, \rho) = \operatorname{Tr} \sigma (\log \sigma - \log \rho)$, which corresponds to the operator convex function $f(x) = x \log x$. Another example is given by $S_s(\sigma, \rho) = 1 - \operatorname{Tr} \sigma^s \rho^{1-s}$, this corresponds to the function $f_s(x) = 1 - x^s$, which is operator convex for $s \in [0, 1]$.

Let $T^* : \mathcal{B} \to \mathcal{A}$ be a unital Schwarz map. Then any $f$-divergence is monotone under $T$ [13], in the sense that

$$S_f(T(\sigma), T(\rho)) \leq S_f(\sigma, \rho).$$

**Theorem 3 ([13]).** *Under the Conditions* 1 *and* 2, *the following are equivalent.*

(i) $T$ *is reversible for* $\mathcal{S}$.
(ii) $S(T(\sigma), T(\rho)) = S(\sigma, \rho)$ *for all* $\sigma \in \mathcal{S}$.
(iii) $T^*(T(\sigma)^{it} T(\rho)^{-it}) = \sigma^{it} \rho^{-it}$ *for* $\sigma \in \mathcal{S}$, $t \in \mathbb{R}$.
(iv) $\operatorname{Tr} T(\sigma)^s T(\rho)^{1-s} = \operatorname{Tr} \sigma^s \rho^{1-s}$ *for all* $\sigma \in \mathcal{S}$ *and some* $s \in (0, 1)$.
(v) $S_f(T(\sigma), T(\rho)) = S_f(\sigma, \rho)$ *for all* $\sigma \in \mathcal{S}$ *and some operator convex function* $f$ *with* $|\operatorname{supp} \mu_f| \geq \dim(\mathcal{H})^2 + \dim(\mathcal{K})^2$, *where* $|X|$ *denotes the number of elements in the set* $X$.
(vi) *Equality holds in* (v) *for all operator convex functions.*
(vii) *Equality holds in* (iv) *for all* $s \in [0, 1]$.
(viii) $T^*_\rho \circ T(\sigma) = \sigma$ *for all* $\sigma \in \mathcal{S}$.

**Remark 1.** The equivalence of (i)–(iii) and (viii) was first proved in [27], for the case when all states are faithful and $T$ is the restriction to a subalgebra, and subsequently for any unital 2-positive map in [29], in the more general setting of von Neumann algebras, see also [15, 16], where Conditions (iv) and (vii) were proved.

The following example shows that, unlike the classical case, preservation of an $f$-divergence with strictly operator convex $f$ is in general not sufficient for reversibility. This solves another open problem of [13], showing that the support condition in Theorem 3(v) cannot be completely removed.

**Example 1.** The function $f(x) = (1+x)^{-1}$, $x \geq 0$ is operator convex and the corresponding measure $\mu_f$ is concentrated in the point $t = 1$, $\mu_f(\{1\}) = 1$. We have

$$S_f(\sigma, \rho) = \operatorname{Tr} \rho (L_\sigma + R_\rho)^{-1}(\rho).$$

We will show that the equality $S_f(T(\sigma), T(\rho)) = S_f(\sigma, \rho)$ does not imply reversibility of $T$.

Let $\mathcal{A}$ be a matrix algebra and let $\sigma \in \mathcal{A}$ be an invertible density matrix. Let $p \in \mathcal{A}$ be a projection such that $\sigma p \neq p\sigma$ and $\operatorname{Tr} p\sigma = \lambda \neq 1/2$. Let $\mathcal{B} \subset \mathcal{A}$ be the abelian subalgebra generated by $p$ and let $T : \mathcal{A} \to \mathcal{B}$ be the trace preserving conditional expectation, then $T(\sigma)$ is the density of the restriction of $\sigma$ to $\mathcal{B}$. Put $x := (1-\lambda)p + \lambda(I - p) \in \mathcal{B}$ and $\rho := (I - x)^{-1}\sigma x$. Then

$$\rho = c^{-1} x \sigma x \geq 0$$

where $c = \lambda(1-\lambda)$, and

$$\operatorname{Tr} \rho = c^{-1} \operatorname{Tr} \sigma x^2 = 1$$

so that $\rho$ is an invertible density matrix as well. Moreover, we also have $T(\rho) = (I - x)^{-1} T(\sigma) x$, so that

$$(L_\sigma + R_\rho)^{-1}(\rho) = x = (L_{T(\sigma)} + R_{T(\rho)})^{-1}(T(\rho))$$

and the equality $S_f(T(\sigma), T(\rho)) = S_f(\sigma, \rho)$ holds. On the other hand, we have from Theorem 5(iv) below that $T$ is reversible if and only if $\sigma = \sigma^A \rho^B$ and $\rho = \rho^A \rho^B$ for some $\sigma^A, \rho^A \in \mathcal{B}^+$ and $\rho^B \in \mathcal{A}^+$. It follows that both $\sigma^A$ and $\rho^A$ commute with $\rho^B$ and, since $\mathcal{B}$ is abelian, this implies that $\sigma^A$ and $\rho^A$ commute with $\sigma$. But this is possible only if $\rho^A$ and $\sigma^A$ are constants. It follows that we must have $\sigma = \rho$ and it is easy to see that this implies that $\sigma$ commutes with $x$, which is not possible by the construction of $x$.

### 3.2. *The commutant Radon–Nikodym derivative*

Let $\rho$, $\sigma$ be density operators in $\mathcal{A}$ and suppose that $\operatorname{supp} \sigma \leq \operatorname{supp} \rho =: p$. The commutant Radon–Nikodym derivative of $\sigma$ with respect to $\rho$ is defined by

$$d(\sigma, \rho) = \rho^{-1/2} \sigma \rho^{-1/2}.$$

Then $d = d(\sigma, \rho)$ is the unique element in $p\mathcal{A}p$, satisfying

$$\operatorname{Tr} \sigma a = \langle I, a \rangle_\sigma = \langle d, a \rangle_\rho. \tag{6}$$

Moreover, $d \geq 0$ and $\|d\|$ is the smallest number $\lambda$ satisfying $\sigma \leq \lambda \rho$, note that $\|d\| \geq 1$ and $\|d\| = 1$ if and only if $\rho = \sigma$.

**Lemma 5.** *Let* $\operatorname{supp} \sigma \leq \operatorname{supp} \rho$. *Let* $T : \mathcal{A} \to \mathcal{B}$ *be a trace preserving positive map. Then*

$$d(T(\sigma), T(\rho)) = T_\rho(d(\sigma, \rho)).$$

**Proof.** Directly by definition of $T_\rho$ and $d(\sigma, \rho)$. □

The following simple lemma provides a useful tool for the analysis of reversibility. Note also that it gives a reversibility condition also for the case when both $T$ and the reverse map $S$ are only required to be positive and trace preserving.

**Lemma 6.** *Let* $\rho$ *be invertible and let* $T : \mathcal{A} \to \mathcal{B}$ *be a trace preserving positive map. Then* $T^*_\rho \circ T(\sigma) = \sigma$ *if and only if* $T^*(d(T(\sigma), T(\rho))) = d(\sigma, \rho)$.

**Proof.** For $a \in \mathcal{A}$, we have by (3) and (6) that

$$\begin{aligned}\langle T^*(d(T(\sigma), T(\rho))), a \rangle_\rho &= \langle d(T(\sigma), T(\rho)), T_\rho(a) \rangle_{T(\rho)} = \operatorname{Tr} T_\rho(a) T(\sigma) \\ &= \operatorname{Tr} a T^*_\rho \circ T(\sigma).\end{aligned}$$

It follows that $T^*(d(T(\sigma), T(\rho))) = d(\sigma, \rho)$ if and only if $\operatorname{Tr} a T^*_\rho \circ T(\sigma) = \operatorname{Tr} a\sigma$ for all $a \in \mathcal{A}$. □

Now we are able to characterize reversibility in terms of the Radon–Nikodym derivative. While (ii) or (iii) give easy conditions for reversibility, Condition (iv) will be necessary for the proof of Theorem 6 below. The last two conditions are not really new, but will be useful in proving Theorem 7.

**Theorem 4.** *Suppose the Conditions* 1 *and* 2 *hold. Let us denote* $\Phi = T^* \circ T_\rho$. *Then the following are equivalent.*

(i) $T$ *is reversible for* $\mathcal{S}$.
(ii) $T^*(d(T(\sigma), T(\rho))) = d(\sigma, \rho)$, *for all* $\sigma \in \mathcal{S}$.
(iii) $d(\sigma, \rho) \in \mathcal{F}_\Phi$, *for all* $\sigma \in \mathcal{S}$.
(iv) $\rho^{it} d(\sigma, \rho) \rho^{-it} \in T^*(\mathcal{M}_{T^*})$, *for all* $\sigma \in \mathcal{S}$ *and* $t \in \mathbb{R}$.
(v) *There is a trace preserving completely positive map* $\hat{S} : \mathcal{B} \to \mathcal{A}$, *such that* $\hat{S} \circ T(\sigma) = \sigma$, $\sigma \in \mathcal{S}$.
(vi) *There are trace preserving completely positive maps* $\hat{T} : \mathcal{A} \to \mathcal{B}$ *and* $\hat{S} : \mathcal{B} \to \mathcal{A}$, *such that* $\hat{T}(\sigma) = T(\sigma)$, $\hat{S} \circ T(\sigma) = \sigma$, $\sigma \in \mathcal{S}$.

**Proof.** By Lemma 6, (ii) is equivalent to $T^*_\rho \circ T(\sigma) = \sigma$ for $\sigma \in \mathcal{S}$, which is equivalent to (i) by Theorem 3(viii). (iii) is the same as (ii), by Lemma 5. Since by Theorem 2(iii), $\mathcal{F}_\Phi$ is a subalgebra in $T^*(\mathcal{M}_{T^*})$ and $\rho^{it}\mathcal{F}_\Phi\rho^{-it} \subseteq \mathcal{F}_\Phi$ for all $t \in \mathbb{R}$, (iii) implies (iv).

Suppose (iv) and let $\mathcal{A}_1$ be the subalgebra generated by $\{\rho^{it}d(\sigma,\rho)\rho^{-it},\ t \in \mathbb{R},\ \sigma \in \mathcal{S}\}$. Then $\mathcal{A}_1 \subseteq T^*(\mathcal{M}_{T^*})$. Let $E : \mathcal{A} \to \mathcal{A}_1$ be the trace preserving conditional expectation. Then its adjoint is the embedding $E^* : \mathcal{A}_1 \hookrightarrow \mathcal{A}$ and since $\rho^{it}\mathcal{A}_1\rho^{-it} \subseteq \mathcal{A}_1$ for all $t \in \mathbb{R}$, the map $E_\rho$ is the $\rho$-preserving conditional expectation, [1]. Hence

$$E^*(d(E(\sigma), E(\rho))) = d(E(\sigma), E(\rho)) = E_\rho(d(\sigma,\rho)) = d(\sigma,\rho).$$

By the equivalence of (ii) and (i) and Theorem 3(viii) (for the map $E$), $E^*_\rho \circ E(\sigma) = \sigma$ for all $\sigma \in \mathcal{S}$.

Let $F^*$ denote the embedding $\mathcal{M}_{T^*} \hookrightarrow \mathcal{B}$, then, as above, its adjoint $F = F^{**} : \mathcal{B} \to \mathcal{M}_{T^*}$ is the trace preserving conditional expectation. Let us define the map $\bar{T} : \mathcal{M}_{T^*} \to T^*(\mathcal{M}_{T^*})$ by $\bar{T} := T^* \circ F^*$. Then since $T^*$ is faithful by Lemma 1, $\bar{T}$ is injective, so that $\bar{T}$ is a *-isomorphism and there is an inverse map $R = (\bar{T})^{-1} : T^*(\mathcal{M}_{T^*}) \to \mathcal{M}_{T^*}$. Define the map $\hat{S} : \mathcal{B} \to \mathcal{A}$ by $\hat{S} := E^*_\rho \circ R^* \circ F$. Then $\hat{S}$ is completely positive and trace preserving. Moreover, $T^* \circ \hat{S}^* = T^* \circ F^* \circ R \circ E_\rho = \bar{T} \circ R \circ E_\rho = E^* \circ E_\rho$, so that $\hat{S} \circ T(\sigma) = (E^* \circ E_\rho)^*(\sigma) = \sigma$ and (v) holds.

Suppose (v). Let $\mathcal{S}_0 := T(\mathcal{S})$ and let $\sigma_0 = T(\sigma)$ for $\sigma \in \mathcal{S}$. Then since $\hat{S}(\sigma_0) = \sigma$ and $T \circ \hat{S}(\sigma_0) = T(\sigma) = \sigma_0$, the map $\hat{S}$ is reversible for $\mathcal{S}_0$. Hence by Theorem 3(viii), the map $\hat{T} := \hat{S}^*_{\rho_0}$ is completely positive and satisfies $\hat{T}(\sigma) = \sigma_0$, this proves (vi). The implication (vi) $\to$ (i) is clear. □

**Remark 2.** Note that by the proof of (v), the completely positive maps $\hat{T}$ and $\hat{S}$ can always be given as adjoints of a composition of a conditional expectation and a *-isomorphism.

**Corollary 1.** *Under the Conditions* 1 *and* 2, $T$ *is reversible for* $\mathcal{S}$ *if and only if* $T$ *is reversible for* $\tilde{\mathcal{S}} := \bigcup\{\rho^{is}\mathcal{S}\rho^{-is},\ s \in \mathbb{R}\}$.

**Proof.** Suppose $T$ is reversible for $\mathcal{S}$. Let $\sigma \in \mathcal{S}$ and let $d = d(\sigma,\rho)$. Then $d \in \mathcal{F}_\Phi$ and therefore also $d(\rho^{is}\sigma\rho^{-is}, \rho) = \rho^{is}d\rho^{-is} \in \mathcal{F}_\Phi$, for all $s \in \mathbb{R}$. □

### 3.3. *Factorization*

In this section, we give a characterization of reversibility in terms of the structure of states in $\mathcal{S}$. More precisely, we show that the elements in $\mathcal{S}$ must have the form of a product of two positive operators, such that $T^*$ is multiplicative on one of them and the other does not depend on $\sigma$. This can be viewed as a quantum version of the classical factorization theorem for sufficient statistics, see, e.g., [33]. The first such factorization result was proved in [22], see also [13, Theorem 6.1]. Similar conditions for the infinite dimensional case are proved in [15, Theorem 6].

**Theorem 5.** *Assume Conditions* 1 *and* 2. *Let* $\Phi = T^* \circ T_\rho$ *and* $\tilde{\Phi} = T_\rho \circ T^*$. *Then the following are equivalent.*

(i) $T$ *is reversible for* $\mathcal{S}$.
(ii) *There is a positive invertible element* $\rho^B \in \mathcal{F}'_\Phi \cap \mathcal{A}$, *such that for each* $\sigma \in \mathcal{S}$,

$$\sigma = T^*(\sigma_0^A)\rho^B, \quad T(\sigma) = \sigma_0^A T(\rho^B),$$

*with some* $\sigma_0^A \in \mathcal{F}^+_{\tilde{\Phi}}$.
(iii) *There is an element* $\rho^B \in \mathcal{A}^+$, *such that for each* $\sigma \in \mathcal{S}$,

$$\sigma = T^*(\sigma_0^A)\rho^B, \quad T(\sigma) = \sigma_0^A T(\rho^B),$$

*with some* $\sigma_0^A \in \mathcal{B}^+$.
(iv) *There is an element* $\rho^B \in \mathcal{A}^+$, *such that each* $\sigma \in \mathcal{S}$ *has the form*

$$\sigma = \sigma^A \rho^B,$$

*where* $\sigma^A$ *is a positive element in* $T^*(\mathcal{M}_{T^*})$.

**Proof.** Let us denote $\sigma_0 := T(\sigma)$ for $\sigma \in \mathcal{S}$. Suppose (i) and let

$$\rho = T^*(\rho_0^A)\rho^B, \quad \rho_0 = \rho_0^A T(\rho^B)$$

be the decomposition from Theorem 2(vi). Then, by Theorems 2 and 4, we have for $\sigma \in \mathcal{S}$,

$$\begin{aligned}\sigma &= \rho^{1/2} d(\sigma,\rho)\rho^{1/2} = \rho^{1/2} T^*(d(\sigma_0,\rho_0))\rho^{1/2} \\ &= T^*(\rho_0^A)^{1/2} T^*(d(\sigma_0,\rho_0)) T^*(\rho_0^A)^{1/2}\rho^B \\ &= T^*((\rho_0^A)^{1/2} d(\sigma_0,\rho_0)(\rho_0^A)^{1/2})\rho^B = T^*(\sigma_0^A)\rho^B\end{aligned}$$

where we put $\sigma_0^A := (\rho_0^A)^{1/2} d(\sigma_0,\rho_0)(\rho_0^A)^{1/2}$. Since $d(\sigma_0,\rho_0) = T_\rho(d(\sigma,\rho)) \in \mathcal{F}^+_{\tilde{\Phi}}$, $\sigma_0^A$ is a positive element in $\mathcal{F}_{\tilde{\Phi}}$. Moreover, $\sigma_0^A = T(\rho^B)^{-1/2}\sigma_0 T(\rho^B)^{-1/2}$, hence

$$\sigma_0 = \sigma_0^A T(\rho^B),$$

where we used Theorem 2(v). This proves (ii). It is clear that (ii) implies (iii).

Suppose (iii). Then for $a \in \mathcal{B}$,

$$\operatorname{Tr} a\sigma_0 = \operatorname{Tr} a\sigma_0^A T(\rho^B) = \operatorname{Tr} T^*(a\sigma_0^A)\rho^B.$$

On the other hand,

$$\operatorname{Tr} a\sigma_0 = \operatorname{Tr} T^*(a)\sigma = \operatorname{Tr} T^*(a) T^*(\sigma_0^A)\rho^B.$$

Putting $a = \sigma_0^A$, we obtain

$$\operatorname{Tr} T^*((\sigma_0^A)^2)\rho^B = \operatorname{Tr} T^*(\sigma_0^A)^2\rho^B.$$

Since $\rho$ is invertible, the decomposition implies that $\rho^B$ must be invertible as well, hence by Schwarz inequality, $T^*((\sigma_0^A)^2) = T^*(\sigma_0^A)^2$. This implies that $\sigma_0^A \in \mathcal{M}_{T^*}$, which proves (iv) with $\sigma^A := T^*(\sigma_0^A)$.

Finally, suppose (iv). Let $\sigma \in \mathcal{S}$. Since both $\sigma^A$ and $\rho^B$ are positive and so is their product $\sigma$, they must commute. It follows that

$$w_t := \sigma^{it}\rho^{-it} = (\sigma^A)^{it}(\rho^A)^{-it} \in T^*(\mathcal{M}_{T^*})$$

for all $t \in \mathbb{R}$, where $\rho = \rho^A\rho^B$ is the decomposition for $\rho$. We have $\rho^{is}w_t\rho^{-is} = w_s^* w_{t+s} \in T^*(\mathcal{M}_{T^*})$ for all $t, s \in \mathbb{R}$. By analytic continuation for $t = -i/2$, we get $\rho^{is}\sigma^{1/2}\rho^{-1/2}\rho^{-is} \in T^*(\mathcal{M}_{T^*})$, hence also $\rho^{is}d(\sigma,\rho)\rho^{-is} \in T^*(\mathcal{M}_{T^*})$ for all $s$. By Theorem 4(v), this implies (i). □

The next corollary shows that the recovery map $T_\rho$ does not depend on the choice of $\rho$. For faithful states, this was proved already in [29].

**Corollary 2.** *Suppose the Conditions* 1 *and* 2 *hold. Then $T$ is reversible for $\mathcal{S}$ if and only if $T_\sigma = T_\rho|_{\operatorname{supp}\sigma\mathcal{A}\operatorname{supp}\sigma}$ for all $\sigma \in \mathcal{S}$.*

**Proof.** Let $\sigma \in \mathcal{S}$, $q := \operatorname{supp}\sigma$, $q_0 := \operatorname{supp} T(\sigma)$ and suppose that $T$ is reversible for $\mathcal{S}$. Let us denote $w = \sigma^{1/2}\rho^{-1/2}$, $w_0 = T(\sigma)^{1/2}T(\rho)^{-1/2}$. By Theorem 5(ii) and Theorem 2, we have

$$w_0 = (\sigma_0^A)^{1/2}(\rho_0^A)^{-1/2} \in \mathcal{F}_{\tilde{\Phi}}$$

and

$$w = T^*(w_0) \in \mathcal{F}_\Phi, \quad w_0 = T_\rho(w).$$

Then for $a \in q\mathcal{A}q$,

$$\begin{aligned} T_\sigma(a) &= T(\sigma)^{-1/2}T(\sigma^{1/2}a\sigma^{1/2})T(\sigma)^{-1/2} \\ &= (w_0^{-1})^* T_\rho(w^* a w) w_0^{-1} = (w_0^{-1})^* T_\rho(w)^* T_\rho(a) T_\rho(w) w_0^{-1} \\ &= q_0 T_\rho(a) q_0. \end{aligned}$$

Since $\rho^B$ is invertible, we must have $q_0 = \operatorname{supp}\sigma_0^A \in \mathcal{F}_{\tilde{\Phi}}$ and $q = \operatorname{supp} T^*(\sigma_0^A) = T^*(q_0)$. Hence also $T_\rho(q) = q_0$ and $q_0 T_\rho(a) q_0 = T_\rho(qaq) = T_\rho(a)$.

Conversely, since $T_\rho$ is unital, the equality $T_\sigma = T_\rho|_{q\mathcal{A}q}$ implies that $T_\sigma^* = T_\rho^*|_{q_0\mathcal{B}q_0}$ by Lemma 2, so that $T_\rho^* \circ T(\sigma) = T_\sigma^* \circ T(\sigma) = \sigma$ and $T$ is reversible for $\mathcal{S}$. □

### 3.4. *Quantum hypothesis testing*

Let $\sigma$ and $\rho$ be density operators in $\mathcal{A}$. Let us consider the problem of testing the hypothesis $H_0 = \rho$ against the alternative $H_1 = \sigma$. Any test is represented by an operator $0 \leq M \leq I$, which corresponds to rejecting the hypothesis. Then we have the error probabilities

$$\alpha(M) = \operatorname{Tr}\rho M, \quad \beta(M) = \operatorname{Tr}\sigma(1 - M).$$

For $s \in [0,1)$, we define the Bayes optimal test to be a minimizer of the expression

$$s\alpha(M) + (1-s)\beta(M) = (1-s)(1 - \mathrm{Tr}(\sigma - t\rho)M), \quad t = \frac{s}{1-s}. \tag{7}$$

Then the minimal Bayes error probability is

$$\Pi_s := \min_{0 \le M \le I} \{s\alpha(M) + (1-s)\beta(M)\} = s\alpha(M_{\frac{s}{1-s}}) + (1-s)\beta(M_{\frac{s}{1-s}})$$

where $M_t$ maximizes the expression $\mathrm{Tr}(\sigma - t\rho)M$ over all $0 \le M \le I$. Below we formulate the quantum version of the Neyman–Pearson lemma. The obtained Bayes optimal tests are called the (quantum) NP tests for $(\rho, \sigma)$.

If $a \in \mathcal{A}$ is a self adjoint operator, we denote by $a_+$ the positive part of $a$, that is, $a_+ = \sum_{i, \lambda_i > 0} p_i$, where $a = \sum_i \lambda_i p_i$ is the spectral decomposition of $a$.

**Lemma 7 ([14, 11]).** *For $t \ge 0$, let $P_{t,+} := \mathrm{supp}(\sigma - t\rho)_+$ and let $P_{t,0}$ be the projection onto the kernel of $\sigma - t\rho$. Then $0 \le M_t \le I$ is a Bayes optimal test if and only if*

$$M_t = P_{t,+} + X_t$$

*with $0 \le X_t \le P_{t,0}$. The minimal Bayes error probability is*

$$\Pi_s = \frac{1}{2}(1 - \|(1-s)\sigma - s\rho\|_1).$$

Let now $T : \mathcal{A} \to \mathcal{B}$ be a trace preserving positive map. Let $s \in (0,1)$, $t = s(1-s)^{-1}$ and let $\Pi_s^0$ be the minimal Bayes error probability for testing the hypothesis $H_0 = T(\rho)$ against $H_1 = T(\sigma)$. For $N \in \mathcal{B}$, $0 \le N \le I$, we have

$$\mathrm{Tr}(T(\sigma) - tT(\rho))N = \mathrm{Tr}(\sigma - t\rho)T^*(N) \le \max_{0 \le M \le I} \mathrm{Tr}(\sigma - t\rho)M$$

so that $\Pi_s^0 \ge \Pi_s$, this is equivalent to the fact that

$$\|T(\sigma - t\rho)\|_1 \le \|\sigma - t\rho\|_1. \tag{8}$$

In [17], equality in (8) was investigated for a pair of invertible density operators, in the case when $T$ is the restriction to a subalgebra. If equality holds for all $t \ge 0$, then the subalgebra must contain some Bayes optimal test for all $s \in [0,1]$, such subalgebras are called 2-sufficient. It was shown that in some cases, 2-sufficiency is equivalent to sufficiency, that is, reversibility of $T$ for $\{\sigma, \rho\}$. From another point of view, this condition was studied also in [5] and it was shown that for a completely positive trace preserving map, the equality implies reversibility for certain sets $\mathcal{S}$.

Since the $L_1$-norm is one of the basic distance measures on states, equivalence between equality in (8) and reversibility is an important open question. We will show below (Theorem 6) that this equivalence holds if equality in (8) is required for all $\sigma$ in the extended family $\tilde{\mathcal{S}} = \bigcup\{\rho^{is}\mathcal{S}\rho^{-is},\ s \in \mathbb{R}\}$. Moreover, Theorem 7 shows this equivalence if equality in (8) holds for $n$ copies of the states, for all $n$.

We will suppose below that $\rho$ is invertible.

**Lemma 8 ([17, Lemma 4]).** *$P_{t,0} \neq 0$ if and only if $t$ is an eigenvalue of $d(\sigma, \rho)$. Moreover, the rank of $P_{t,0}$ is equal to the multiplicity of $t$.*

**Lemma 9.** *The function $t \mapsto P_{t,+}$ is right-continuous. Moreover,*

$$\lim_{s \to t^-} P_{s,+} = P_{t,+} + P_{t,0}, \quad t \geq 0.$$

**Proof.** Let $\rho(t) := \sigma - t\rho$ for $t \in \mathbb{R}$. Let $\lambda_1^\downarrow(t), \ldots, \lambda_N^\downarrow(t)$ denote the decreasingly ordered eigenvalues of $\rho(t)$ (with multiplicities). For $t_1, t_2 \in \mathbb{R}$, we have $\rho(t_1) = \rho(t_2) + (t_2 - t_1)\rho$. By Weyl's perturbation theorem [3, Corollary III.2.6], this implies that

$$\max_j |\lambda_j^\downarrow(t_1) - \lambda_j^\downarrow(t_2)| \leq |t_1 - t_2| \|\rho\|.$$

Moreover, since $\rho$ is invertible, we obtain by [3, Corollary III.2.2] that

$$\lambda_j^\downarrow(t_2) < \lambda_j^\downarrow(t_2) + (t_2 - t_1)\lambda_N^\downarrow(\rho) \leq \lambda_j^\downarrow(t_1)$$

when $t_1 < t_2$, where $\lambda_N^\downarrow(\rho)$ denotes the smallest eigenvalue of $\rho$. Hence the functions $t \mapsto \lambda_j^\downarrow(t)$ are continuous and strictly decreasing.

It is clear that for $t < 0$ all $\lambda_j^\downarrow(t)$ are strictly positive, and that $\lambda_j^\downarrow(t) = 0$ for some index $j$ if and only if $P_{t,0} \neq 0$. Let $0 \leq t_1 < \cdots < t_n$ be the eigenvalues of $d(\sigma, \rho)$ and put $t_0 := 0$, $t_{n+1} := \infty$. Then there are indices $i_k \in \{1, \ldots, N\}$, $k = 1, \ldots, n$, such that $N = i_1 > i_2 > \cdots > i_n > i_{n+1} := 0$ and for every $t \in [t_{k-1}, t_k)$ the strictly positive eigenvalues of $\rho(t)$ are given by $\lambda_1^\downarrow(t), \ldots, \lambda_{i_k}^\downarrow(t)$.

Let $t \in [t_{k-1}, t_k)$ and let $\gamma(t)$ be a circle, contained entirely in the open half-plane of complex numbers having strictly positive real parts and enclosing all $\lambda_1^\downarrow(t), \ldots, \lambda_{i_k}^\downarrow(t)$. By continuity of $\lambda_j^\downarrow$, there is some $\delta > 0$ such that $\gamma(t)$ encloses $\lambda_1^\downarrow(s), \ldots, \lambda_{i_k}^\downarrow(s)$ for all $s \in (t - \delta, t + \delta)$ and $[t, t + \delta) \subset [t_{k-1}, t_k)$. Then

$$P_{s,+} = \frac{1}{2i\pi} \oint_{\gamma(t)} (zI - \rho(s))^{-1} dz, \quad s \in [t, t + \delta).$$

This implies that $t \mapsto P_{t,+}$ is right-continuous. Let now $t \in (t_{k-1}, t_k)$, then we can find $\delta > 0$ as above, but such that, moreover, $(t - \delta, t + \delta) \subset (t_{k-1}, t_k)$. In this case,

$$P_{s,+} = \frac{1}{2i\pi} \oint_{\gamma(t)} (zI - \rho(s))^{-1} dz, \quad s \in (t - \delta, t + \delta)$$

so that $t \mapsto P_{t,+}$ is continuous at $t$. Suppose $t = t_{k-1}$, then by definition of $i_k$ and $t_{k-1}$, we must have

$$\lambda_j^\downarrow(t_{k-1}) \begin{cases} > 0 & j \leq i_k, \\ = 0 & j = i_k + 1, \ldots, i_{k-1}, \\ < 0 & j > i_{k-1}. \end{cases}$$

Let $\gamma'_k$ be a circle in the complex plane, enclosing $\lambda_1^\downarrow(t_{k-1}),\dots,\lambda_{i_k}^\downarrow(t_{k-1})$ and 0, but such that the closed disc encircled by $\gamma'_k$ does not contain any other eigenvalue of $\rho(t_{k-1})$. Then there is some $\delta>0$ such that $(t_{k-1}-\delta,t_{k-1})\subset[t_{k-2},t_{k-1})$ and

$$P_{s,+}=\frac{1}{2i\pi}\oint_{\gamma'_k}(zI-\rho(s))^{-1}dz,\quad s\in(t_{k-1}-\delta,t_{k-1}).$$

It follows that $\lim_{s\to t_{k-1}^-}P_{s,+}=P_{t_{k-1},+}+P_{t_{k-1},0}$. Since $P_{t,0}=0$ for $t\notin\{t_1,\dots,t_n\}$, this proves the assertion. □

Let us denote $Q_{t,+}:=\mathrm{supp}(T(\sigma)-tT(\rho))_+$ and $Q_{t,0}$ the projection onto the kernel of $T(\sigma)-tT(\rho)$.

**Lemma 10.** *Let $T:\mathcal{A}\to\mathcal{B}$ be a trace preserving positive map and suppose that both $\rho$ and $T(\rho)$ are invertible. The following are equivalent.*

(i) $\|T(\sigma)-tT(\rho)\|_1=\|\sigma-t\rho\|_1$, *for all* $t\in\mathbb{R}$.
(ii) $P_{t,+}=T^*(Q_{t,+})$, $P_{t,0}=T^*(Q_{t,0})$ *for* $t\in\mathbb{R}$.

**Proof.** Since $Q_{t,+}$ is an NP test for $(T(\rho),T(\sigma))$, (i) implies that

$$\mathrm{Tr}(T(\sigma)-tT(\rho))Q_{t,+}=\mathrm{Tr}(\sigma-t\rho)T^*(Q_{t,+})=\max_{0\le M\le I}\mathrm{Tr}(\sigma-t\rho)M$$

so that $T^*(Q_{t,+})$ is an NP test for $(\rho,\sigma)$. By Lemma 7, there is some $0\le X_t\le P_{t,0}$, such that $T^*(Q_{t,+})=P_{t,+}+X_t$. It follows that $P_{t,+}=T^*(Q_{t,+})$ holds for all $t$ such that $P_{t,0}=0$, that is, for $t\in\mathbb{R}\backslash\{t_1,\dots,t_n\}$. Since $t\mapsto P_{t,+}$ and $t\mapsto T^*(Q_{t,+})$ are right continuous, it follows that $T^*(Q_{t,+})=P_{t,+}$ for all $t$. On the other hand, by Lemma 9 we have for all $t$

$$P_{t,+}+P_{t,0}=\lim_{s\to t^-}P_{s,+}=\lim_{s\to t^-}T^*(Q_{s,+})=T^*(Q_{t,+})+T^*(Q_{t,0})$$

hence $P_{t,0}=T^*(Q_{t,0})$ for all $t$. The converse is obvious. □

**Theorem 6.** *Assume the Conditions* 1 *and* 2. *Then*

(i) *$T$ is reversible for $\mathcal{S}$ if and only if*

$$\|\sigma-t\rho\|_1=\|T(\sigma)-tT(\rho)\|_1,\quad \sigma\in\tilde{\mathcal{S}},\ \ t\ge 0. \tag{9}$$

(ii) *Suppose that $\rho^{is}\mathcal{S}\rho^{-is}\subseteq\mathcal{S}$ for all $s\in\mathbb{R}$. Then $T$ is reversible for $\mathcal{S}$ if and only if*

$$\|\sigma-t\rho\|_1=\|T(\sigma)-tT(\rho)\|_1,\quad \sigma\in\mathcal{S},\ \ t\ge 0. \tag{10}$$

(iii) *Suppose that $\mathcal{B}$ is abelian. Then $T$ is reversible for $\mathcal{S}$ if and only if* (10) *holds. Moreover, in this case all elements in $\mathcal{S}$ commute.*
(iv) *Suppose that all elements in $\mathcal{S}$ commute with $\rho$. Then $T$ is reversible for $\mathcal{S}$ if and only if* (10) *holds.*

**Proof.** (i) By Corollary 1, $T$ is reversible for $\mathcal{S}$ if and only if it is reversible for $\tilde{\mathcal{S}}$. By monotonicity (8), we get (9).

For the converse, let $\sigma \in \tilde{\mathcal{S}}$. Then by Lemma 10, (9) implies that $P_{t,0} = T^*(Q_{t,0})$ for the corresponding projections for $\sigma$ and $\rho$. This implies that $Q_{t,0} \in \mathcal{M}_{T^*}$ and $P_{t,0} \in T^*(\mathcal{M}_{T^*})$.

Let $t_1, \dots, t_n$ be the eigenvalues of $d = d(\sigma, \rho)$ and let $F_1, \dots, F_n$ be the corresponding eigenprojections. Denote $P_i := P_{t_i,0}$. Then we have $(d - t_i)\rho^{1/2} P_i = \rho^{-1/2}(\sigma - t_i\rho)P_i = 0$ and this implies

$$d\rho^{1/2} \sum_i P_i = \rho^{1/2} \sum_i t_i P_i.$$

Moreover, any vector in the range of $\rho^{1/2} P_i \rho^{1/2}$ is an eigenvector of $d$, so that $\operatorname{supp}(\rho^{1/2} P_i \rho^{1/2}) \leq F_i$ and by Lemma 8, $\operatorname{rank}(F_i) = \operatorname{rank}(P_i) = \operatorname{rank}(\rho^{1/2} P_i \rho^{1/2})$. It follows that $\sum_i P_i$ is invertible, so that $d(\sigma, \rho) = \rho^{1/2} c \rho^{-1/2}$, with

$$c := \sum_i t_i P_i \left( \sum_j P_j \right)^{-1} \in T^*(\mathcal{M}_{T^*}).$$

It follows that for $s \in \mathbb{R}$ and $\sigma \in \mathcal{S}$, $\rho^{is-1/2} d(\sigma, \rho) \rho^{1/2-is} \in T^*(\mathcal{M}_{T^*})$. By analytic continuation, we get $\rho^{it} d(\sigma, \rho) \rho^{-it} \in T^*(\mathcal{M}_{T^*})$ for all $t \in \mathbb{R}$, which implies that $T$ is reversible for $\mathcal{S}$, by Theorem 4.

(ii) Clearly follows from (i).

(iii) Let $\sigma \in \mathcal{S}$ and let $P_{t,0}$ and $Q_{t,0}$ be the corresponding projections. Note that since $\mathcal{B}$ is commutative, $Q_{t,0}$ must commute for all $t$. Suppose that (10) holds, then $P_{t,0} = T^*(Q_{t,0})$ and, since then $Q_{t,0} \in \mathcal{M}_{T^*}$, this implies that all $P_{t,0}$ commute as well. As in the proof of (i), $d(\sigma, \rho) = \rho^{1/2} c \rho^{-1/2}$, where we now have $c \geq 0$. This implies that $d(\sigma, \rho)\rho = \rho^{1/2} c \rho^{1/2} \geq 0$, hence $d(\sigma, \rho)\rho = \rho d(\sigma, \rho)$ and therefore also $\sigma\rho = \rho\sigma$. This implies that $\rho^{is}\sigma\rho^{-is} = \sigma$ and the statement follows by (ii). The converse implication is clear.

(iv) Follows from (ii). □

### 3.5. *Quantum Chernoff and Hoeffding distances*

Let $n \in \mathbb{N}$ and suppose we are given $n$ identical copies of the states $\rho^{\otimes n}, \sigma^{\otimes n} \in \mathcal{S}(\mathcal{A}^{\otimes n})$. Consider the problem of testing the hypothesis $H_0 = \rho^{\otimes n}$ against $H_1 = \sigma^{\otimes n}$. Then the minimum Bayes error probability is

$$\Pi_{s,n} = \frac{1}{2}(1 - \|(1-s)\sigma^{\otimes n} - s\rho^{\otimes n}\|_1).$$

It is an important result of [2] that as $n \to \infty$, the probabilities $\Pi_{s,n}$ decay exponentially fast and the rate of convergence is given by

$$\lim_n -\frac{1}{n} \log \Pi_{s,n} = -\log\left( \inf_{0 \leq u \leq 1} \operatorname{Tr} \sigma^u \rho^{1-u} \right) =: C(\sigma, \rho) \tag{11}$$

for any $s \in [0,1]$, where we put $x^0 = \operatorname{supp} x$ for any positive $x \in \mathcal{A}$. The quantity $C(\sigma,\rho)$ is called the quantum Chernoff distance. Note that $C$ is related to the convex quantum $f$-divergence $S_u(\sigma,\rho)$, but one can show that $C$ itself is not an $f$-divergence [13]. Nevertheless, if $T$ is the adjoint of a unital Schwarz map, then $C$ satisfies monotonicity:

$$C(\sigma,\rho) \geq C(T(\sigma),T(\rho))$$

and, moreover, $C(\sigma,\rho)=0$ if and only if $\sigma=\rho$.

Let us consider again the problem of testing the hypothesis $H_0=\rho$ against the alternative $H_1=\sigma$. Let $0 \leq M \leq I$ be a test. Differently from the Bayesian approach, in the asymmetric approach the error probability $\alpha(M)$ is bounded, $\alpha(M) \leq \epsilon$ for some fixed $\epsilon > 0$. The error probability $\beta(M)$ is then minimalized over all tests, under this constraint,

$$\beta_\epsilon := \inf\{\beta(M),\, 0 \leq M \leq I,\, \alpha(M) \leq \epsilon\}.$$

Suppose we have $n$ independent copies of the states $\sigma^{\otimes n}$ and $\rho^{\otimes n}$ and let $M_n \in \mathcal{A}^{\otimes n}$. Here we require that the probabilities $\alpha(M_n)$ decay exponentially as $n \to \infty$. Let $r > 0$ and put

$$\beta_{r,n} := \inf\{\beta(M_n),\ 0 \leq M_n \leq I,\ \alpha(M_n) \leq e^{-nr}\}.$$

The following equality was proved in [9, 23]: For $r > 0$,

$$\lim_n -\frac{1}{n}\log\beta_{r,n} = \sup_{0\leq u<1}\frac{-ur-\log\operatorname{Tr}\rho^u\sigma^{1-u}}{1-u} =: H_r(\rho,\sigma).$$

The limit expression is called the quantum Hoeffding distance. Similarly as the Chernoff distance, $H_r$ is not an $f$-divergence [13], but it is related to $S_u$. This implies the monotonicity

$$H_r(T(\sigma),T(\rho)) \leq H_r(\sigma,\rho)$$

for $T$ the adjoint of a unital Schwarz map. Moreover, by [12], see also [13],

$$H_0(\sigma,\rho) := \lim_{r\to 0} H_r(\sigma,\rho) = S(\sigma,\rho) = \operatorname{Tr}\sigma(\log\sigma - \log\rho)$$

holds if $\operatorname{supp}\sigma \leq \operatorname{supp}\rho$.

Suppose that $q := \operatorname{supp}\sigma \leq \operatorname{supp}\rho$, then the function $[0,\infty) \ni r \mapsto H_r(\sigma,\rho)$ has the following properties [12], see also [2]:

The function is convex and lower semicontinuous, for $r \in [0, S_\sigma(\rho,\sigma)]$ it is strictly convex and decreasing, and for $r \geq S_\sigma(\rho,\sigma)$ it has a constant value $H_r(\sigma,\rho) = -\log\operatorname{Tr} q\rho$, here

$$S_\sigma(\rho,\sigma) = -\log\operatorname{Tr} q\rho + \frac{1}{\operatorname{Tr} q\rho}\operatorname{Tr}\rho(\log\rho - \log\sigma)q.$$

Note that if $\operatorname{supp}\sigma = \operatorname{supp}\rho$, then $S_\sigma(\rho,\sigma) = S(\rho,\sigma)$.

**Proposition 3 ([13]).** *Let $\sigma$ and $\rho$ be two density operators in $\mathcal{A}$ such that* $\operatorname{supp}\sigma = \operatorname{supp}\rho$. *Let $T : \mathcal{A} \to \mathcal{B}$ be the adjoint of a unital Schwarz map and suppose that one of the following conditions holds*:

(i) $C(\sigma, \rho) = C(T(\sigma), T(\rho))$.
(ii) $H_r(\sigma, \rho) = H_r(T(\sigma), T(\rho))$ *for some* $r \in [0, S(T(\rho), T(\sigma))]$.

*Then* $T^*_\rho \circ T(\sigma) = \sigma$.

**Theorem 7.** *Assume the Conditions* 1 *and* 2. *Then the following are equivalent.*

(i) $T$ *is reversible for* $\mathcal{S}$.
(ii) $C(\sigma, \rho) = C(T(\sigma), T(\rho))$ *for all* $\sigma \in co(\mathcal{S})$.
(iii) $\|\sigma^{\otimes n} - t\rho^{\otimes n}\|_1 = \|T(\sigma)^{\otimes n} - tT(\rho)^{\otimes n}\|_1$ *for all* $\sigma \in \mathcal{S}$, $t \geq 0$ *and* $n \in \mathbb{N}$.
(iv) $H_r(\sigma, \rho) = H_r(T(\sigma), T(\rho))$ *for all* $\sigma \in \mathcal{S}$ *and* $r \geq 0$.

*Suppose moreover that there is some* $\mathcal{S}_0 \subset \mathcal{S}$, *such that* $\mathcal{S} \subseteq \overline{co}(\mathcal{S}_0 \cup \{\rho\})$ *and* $T(\rho) \notin \overline{T(\mathcal{S}_0)}$. *Then there exists some* $r_0 > 0$ *such that* (i)–(iv) *are equivalent to*

(v) $H_r(\sigma, \rho) = H_r(T(\sigma), T(\rho))$ *for all* $\sigma \in co(\mathcal{S})$ *and some* $r \in [0, r_0]$.

**Proof.** Since $T$ is reversible for $\mathcal{S}$ if and only if it is reversible for $co(\mathcal{S})$, (i) implies (ii) by monotonicity of $C$. Conversely, suppose (ii) and let $\sigma \in \mathcal{S}$, then $\sigma_1 := \frac{1}{2}(\sigma+\rho)$ is an invertible element in $co(\mathcal{S})$. Proposition 3 now implies that $T^*_\rho \circ T(\sigma_1) = \sigma_1$ and by (4), we have also $T^*_\rho \circ T(\sigma) = \sigma$.

Further, suppose (i), then by Theorem 4(vi), there are trace preserving completely positive maps $\hat{T} : \mathcal{A} \to \mathcal{B}$ and $\hat{S} : \mathcal{B} \to \mathcal{A}$, such that $\hat{T}(\sigma) = T(\sigma)$, $\hat{S} \circ T(\sigma) = \sigma$, $\sigma \in \mathcal{S}$. It follows that $T(\sigma)^{\otimes n} = \hat{T}(\sigma)^{\otimes n} = \hat{T}^{\otimes n}(\sigma^{\otimes n})$ and $\sigma^{\otimes n} = \hat{S}^{\otimes n}(T(\sigma)^{\otimes n})$, for all $\sigma \in \mathcal{S}$, where $\hat{T}^{\otimes n}$ and $\hat{S}^{\otimes n}$ are completely positive and trace preserving. By monotonicity of the $L_1$-norm, this implies (iii). The implications (iii) $\Rightarrow$ (iv) $\Rightarrow$ (i) were already proved in [13].

Suppose now that the additional condition holds. Let us choose some $\varepsilon \in (0, 1)$ and put

$$r_0 := \inf_{\sigma \in \mathcal{S}_0} S(T(\rho), T(\varepsilon\rho + (1-\varepsilon)\sigma)).$$

Then if $r_0 = 0$, there exists a sequence $\sigma_n \in \mathcal{S}_0$, such that $S(T(\rho), T(\varepsilon\rho + (1-\varepsilon)\sigma_n)) \to 0$. This implies that $T(\sigma_n) \to T(\rho)$, so that $T(\rho) \in \overline{T(\mathcal{S}_0)}$, which is not possible. Hence $r_0 > 0$.

Suppose (v) holds and let $\sigma \in \mathcal{S}_0$. Denote $\sigma_\varepsilon = \varepsilon\rho + (1-\varepsilon)\sigma$. Then $0 \leq r \leq S(T(\rho), T(\sigma_\varepsilon))$. Since $\sigma_\varepsilon$ is invertible, we can apply Proposition 3, which implies that $T^*_\rho \circ T(\sigma_\varepsilon) = \sigma_\varepsilon$ and therefore also $T^*_\rho \circ T(\sigma) = \sigma$ for all $\sigma \in \mathcal{S}_0$. Since $\mathcal{S} \subseteq \overline{co}(\mathcal{S}_0 \cup \{\rho\})$, this implies (i). The implication (i) $\Rightarrow$ (v) follows by monotonicity. $\square$

**Remark 3.** Note that if all elements in $\mathcal{S}$ are invertible, then we may replace $co(\mathcal{S})$ by $\mathcal{S}$ in (ii) and by $\mathcal{S}_0$ in (v), where we put $r_0 := \inf_{\sigma \in \mathcal{S}_0} S(T(\rho), T(\sigma))$.

### 3.6. *Quantum Fisher information and $\chi^2$-divergence*

Let us denote by $\mathcal{D}$ the set of invertible density operators in $\mathcal{A}$. Then $\mathcal{D}$ is a differentiable manifold, where the tangent space at each point $\rho \in \mathcal{D}$ is the vector space $\mathcal{T}_\rho$ of traceless self-adjoint elements in $\mathcal{A}$.

A monotone metric on $\mathcal{D}$ is a Riemannian metric $\lambda_\rho$, satisfying

$$\lambda_\rho(x,x) \geq \lambda_{T(\rho)}(T(x),T(x)), \quad x \in \mathcal{T}_\rho, \quad \rho \in \mathcal{D} \tag{12}$$

for any completely positive trace preserving map $T : \mathcal{A} \to \mathcal{B}$.

It was proved by Petz in [30] that any monotone metric has the form

$$\lambda_\rho(x,y) = \mathrm{Tr}(J_\rho^f)^{-1}(x)y$$

with $J_\rho^f = f(\Delta_\rho)R_\rho$, where $\Delta_\rho := \Delta_{\rho,\rho} = L_\rho R_\rho^{-1}$, and $f : (0,\infty) \to (0,\infty)$ an operator monotone function satisfying the symmetry $f(t) = tf(t^{-1})$. Under the normalization condition $f(1) = 1$, the restriction of $\lambda_\rho$ to the submanifold of diagonal elements in $\mathcal{D}$ coincides with the classical Fisher information for probability measures on a finite set, moreover, the monotonicity condition (12) characterizes the classical Fisher information up to multiplication by a constant. Accordingly, any monotone metric with the above normalization is called a quantum Fisher information.

The operator $J_\rho^f$ satisfies [28, 24]

$$J_{T(\rho)}^f \geq T J_\rho^f T^*$$

for any operator monotone (not necessarily symmetric or normalized) function $f$ and $T : \mathcal{A} \to \mathcal{B}$ the adjoint of a unital Schwarz map. This is equivalent to [30]

$$(J_\rho^f)^{-1} \geq T^*(J_{T(\rho)}^f)^{-1}T, \tag{13}$$

which implies that the monotonicity (12) holds for all such $f$ and $T$.

A related quantity is the quantum version of the $\chi^2$-divergence, which was introduced in [32] as

$$\chi^2_{1/f}(\sigma,\rho) = \lambda_\rho^f(\sigma-\rho,\sigma-\rho)$$

where $\lambda_\rho^f$ is a monotone metric.

Let now $f : (0,\infty) \to (0,\infty)$ be operator monotone. Then $t \mapsto f(t)^{-1}$ is a non-negative operator monotone decreasing function on $(0,\infty)$. By [7], for each such function there is a positive Borel measure $\nu_f$ with support in $[0,\infty)$ and $\int_0^\infty (1 + s^2)^{-1} d\nu_f(s) < \infty$, $\int_0^\infty s(1+s^2)^{-1} d\nu_f(s) < \infty$, such that

$$f(t)^{-1} = \int_0^\infty \frac{1}{s+t} d\nu_f(s) = \int_0^\infty f_s(t)^{-1} d\nu_f(s)$$

where $f_s(t) = s + t$, $t \in \mathbb{R}^+$. Then it follows that

$$(J^f_\rho)^{-1} = f(L_\rho R_\rho^{-1})^{-1} R_\rho^{-1} = \int_0^\infty (sR_\rho + L_\rho)^{-1} d\nu_f(s) = \int_0^\infty (J^s_\rho)^{-1} d\nu_f(s) \quad (14)$$

where $J^s_\rho := J^{f_s}_\rho = sR_\rho + L_\rho$.

**Lemma 11.** *Let $T : \mathcal{A} \to \mathcal{B}$ be the adjoint of a unital Schwarz map. Let $x \in \mathcal{A}$. Then for $s \geq 0$,*

$$\operatorname{Tr} x^*(sR_\rho + L_\rho)^{-1}(x) \geq \operatorname{Tr} T(x)^*(sR_{T(\rho)} + L_{T(\rho)})^{-1}(T(x)) \quad (15)$$

*and equality holds if and only if*

$$(sR_\rho + L_\rho)^{-1}(x) = T^*[(sR_{T(\rho)} + L_{T(\rho)})^{-1}(T(x))]. \quad (16)$$

**Proof.** Since the function $f_s$ is operator monotone, the inequality (15) follows from (13) for $f = f_s$. If equality holds for some $x \in \mathcal{A}$, then

$$\langle x, ((J^s_\rho)^{-1} - T^*(J^s_{T(\rho)})^{-1}T)(x)\rangle = 0$$

which again by (13) is equivalent to $((J^s_\rho)^{-1} - T^*(J^s_{T(\rho)})^{-1}T)(x) = 0$. □

It follows from the above lemma and the integral representation (14) that $\lambda^f_\rho(x,x) = \lambda^f_{T(\rho)}(T(x), T(x))$ if and only if (16) holds for all $s \in \operatorname{supp} \nu_f$, that is,

$$(s + \Delta_\rho)^{-1}(x\rho^{-1}) = T^*[(s + \Delta_{T(\rho)})^{-1}(T(x)T(\rho)^{-1})], \quad s \in \operatorname{supp} \nu_f. \quad (17)$$

Let now $x \in \mathcal{T}_\rho$. Then since $\rho$ is invertible, there exists some interval $I \ni 0$ such that $\sigma_u := \rho + ux \in \mathcal{S}(\mathcal{A})$ for $u \in I$. Let us denote by $I_{\rho,x}$ the largest such interval and let $\mathcal{S}_{\rho,x} := \{\sigma_u, u \in I_{\rho,x}\}$.

**Proposition 4.** *Let $\rho \in \mathcal{D}$, $x \in \mathcal{T}_\rho$ and $T : \mathcal{A} \to \mathcal{B}$ be such that $T$ and $\mathcal{S}_{\rho,x}$ satisfy the Conditions* 1 *and* 2. *Then the following are equivalent.*

- (i) $\lambda^f_\rho(x,x) = \lambda^f_{T(\rho)}(T(x), T(x))$ *for a monotone metric such that* $|\operatorname{supp} \nu_f| \geq |\operatorname{spec}(\Delta_\rho) \cup \operatorname{spec}(\Delta_{T(\rho)})|$.
- (ii) $\rho^{it} x \rho^{-it-1} = T^*(T(\rho)^{it} T(x) T(\rho)^{-it-1})$, $t \in \mathbb{R}$.
- (iii) $\rho^{-1/2} x \rho^{-1/2} = T^*(T(\rho)^{-1/2} T(x) T(\rho)^{-1/2})$.
- (iv) $T$ *is reversible for* $\mathcal{S}_{\rho,x}$.
- (v) $\lambda^f_\rho(x,x) = \lambda^f_{T(\rho)}(T(x), T(x))$ *for any monotone metric* $\lambda^f_\rho$.

**Proof.** Note that by the assumptions, $T(\rho)$ must be invertible. Suppose (i), then (17) holds and by [13, Lemma 5.2], this implies that

$$h(\Delta_\rho) x \rho^{-1} = h(\Delta_{T(\rho)}) T(x) T(\rho)^{-1}$$

for any complex-valued function $h$ on $\mathrm{spec}(\Delta_\rho) \cup \mathrm{spec}(\Delta_{T(\rho)})$. In particular, for $h(\lambda) = \lambda^{it}$, we get (ii). We have (ii) $\Rightarrow$ (iii) by analytic continuation for $t = i/2$. Suppose (iii) and let $\sigma_u \in \mathcal{S}_{\rho,x}$. Then

$$\begin{aligned} T^*(d(T(\sigma_u), T(\rho))) &= T^*(I + uT(\rho)^{-1/2}T(x)T(\rho)^{-1/2}) \\ &= I + u\rho^{-1/2}x\rho^{-1/2} = d(\sigma_u, \rho) \end{aligned}$$

and by Theorem 4, this implies (iv). (iv) implies (v) by monotonicity of Fisher information. The implication (v) $\Rightarrow$ (i) is trivial. □

Let $\mathcal{S} \subset \mathcal{S}(\mathcal{A})$ and let $\mathrm{Lin}(\mathcal{S}) = \mathrm{span}\,\{\sigma_1 - \sigma_2 : \sigma_1, \sigma_2 \in \mathcal{S}\}$. Then $\mathrm{Lin}(\mathcal{S})$ is a vector subspace in the real vector space of self-adjoint traceless operators.

**Theorem 8.** *Suppose that the Conditions* 1 *and* 2 *hold. Then the following are equivalent.*

(i) *$T$ is reversible for $\mathcal{S}$.*
(ii) *$\lambda^f_\rho(x, x) = \lambda^f_{T(\rho)}(T(x), T(x))$ for all $x \in \mathrm{Lin}(\mathcal{S})$ and all monotone metrics.*
(iii) *$\chi^2_{1/f}(\sigma, \rho) = \chi^2_{1/f}(T(\sigma), T(\rho))$ for all $\sigma \in \mathcal{S}$ and all $\chi^2$-divergences.*
(iv) *The equality in* (ii) *holds for some symmetric positive operator monotone function $f$ such that $|\mathrm{supp}\,\mu_f| \geq \dim(\mathcal{H})^2 + \dim(\mathcal{K})^2$.*
(v) *The equality in* (iii) *holds for some $f$ as in* (iv).

**Proof.** (i) implies (ii) by monotonicity of Fisher information and the implication (ii) $\Rightarrow$ (iii) is clear. We also have (ii) $\Rightarrow$ (iv) and both (iv) and (iii) imply (v). It is therefore enough to prove (v) $\Rightarrow$ (i). So suppose (v) and let $\sigma \in \mathcal{S}$. Put $x = \sigma - \rho$ in Proposition 4(iii), then it follows that $T^*(d(T(\sigma), T(\rho))) = d(\sigma, \rho)$ for $\sigma \in \mathcal{S}$ which implies (i) by Theorem 4. □

**Remark 4.** An important example of a quantum Fisher information, respectively $\chi^2$-divergence, is given by $f(t) = \frac{1}{2}f_1(t) = \frac{1}{2}(1+t)$. In this case, $\nu_f$ is concentrated in $t = 1$ and $\lambda^f_\rho(x, y) = 2\mathrm{Tr}\, y(L_\rho + R_\rho)^{-1}(x)$ is called the Bures metric. It is the smallest element in the family of quantum Fisher informations. The simple example below shows that preservation of the Bures metric does not imply reversibility, so that, once again, the support condition in (iv) respectively (v) of the above theorem cannot be dropped.

So let $y = y^* \in \mathcal{A}$ be such that $\rho y \neq y\rho$ and $\mathrm{Tr}\,\rho y = 0$, and let $\mathcal{C} \subset \mathcal{A}$ be the commutative subalgebra generated by $y$. Then $z := \rho y + y\rho \in \mathcal{T}_\rho$ and, by replacing $y$ by $ty$ for some $t > 0$ if necessary, we may suppose that $\sigma := \rho + z \in \mathcal{D}$. Let $T : \mathcal{A} \to \mathcal{C}$ be the trace preserving conditional expectation, then $T(\sigma) = T(\rho) + T(z) = T(\rho) + T(\rho)y + yT(\rho)$. This implies that

$$(L_\rho + R_\rho)^{-1}(\sigma - \rho) = y = (L_{T(\rho)} + R_{T(\rho)})^{-1}(T(\sigma) - T(\rho))$$

which implies that $\chi^2_{1/f}(\sigma, \rho) = \chi^2_{1/f}(T(\sigma), T(\rho))$.

On the other hand, if $T$ is reversible, then by Theorem 5(iv), $\rho$ and $\sigma$ must commute. But we have $[\sigma, \rho] = [\rho^2, y] \neq 0$.


## Acknowledgments

I would like to thank Milán Mosonyi for useful comments on the manuscript. I would also like to thank the anonymous referee for a number of valuable suggestions, especially for the proof of Lemma 9. The work was supported by the grants VEGA 2/0032/09 and meta-QUTE ITMS 26240120022.


## References


[1] L. Accardi and C. Cecchini, Conditional expectations in von Neumann algebras and a theorem of Takesaki, *J. Funct. Anal.* **45** (1982) 245–273.

[2] K. M. R. Audenaert, M. Nussbaum, A. Szkola and F. Verstraete, Asymptotic error rates in quantum hypothesis testing, *Comm. Math. Phys.* **279** (2008) 251–283.

[3] R. Bhatia, *Matrix Analysis* (Springer, 1997).

[4] R. Bhatia, *Positive Definite Matrices* (Princeton University Press, 2007).

[5] R. Blume-Kohout, H. K. Ng, D. Poulin and L. Viola, Information preserving structures: A general framework for quantum zero-error information, *Phys. Rev. A* **82** (2010) 062306.

[6] A. Datta, A condition for the nullity of quantum discord (2010); arXiv:1003.5256.

[7] F. Hansen, Trace functions as Laplace transforms, *J. Math. Phys.* **47** (2006) 043504.

[8] M. Hayashi, *Quantum Information: An Introduction* (Springer, 2006).

[9] M. Hayashi, Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding, *Phys. Rev. A* **76** (2007) 062301.

[10] P. Hayden, R. Jozsa, D. Petz and A. Winter, Structure of states which satisfy strong subadditivity of quantum entropy with equality, *Comm. Math. Phys.* **246** (2004) 359–374.

[11] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[12] F. Hiai, M. Mosonyi and T. Ogawa, Error exponents in hypothesis testing for correlated states on a spin chain, *J. Math. Phys.* **49** (2008) 032112.

[13] F. Hiai, M. Mosonyi, D. Petz and C. Beny, Quantum $f$-divergences and error correction, *Rev. Math. Phys.* **23** (2011) 691–747.

[14] A. S. Holevo, On asymptotically optimal hypothesis testing in quantum statistics, *Theory Probab. Appl.* **23** (1978) 411–415.

[15] A. Jenčová and D. Petz, Sufficiency in quantum statistical inference, *Comm. Math. Phys.* **263** (2006) 259276.

[16] A. Jenčová and D. Petz, Sufficiency in quantum statistical inference. A survey with examples, *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **9** (2006) 331–351.

[17] A. Jenčová, Quantum hypothesis testing and sufficient subalgebras, *Lett. Math. Phys.* **93** (2010) 15–27.

[18] A. Jenčová and M. B. Ruskai, A unified treatment of convexity of relative entropy and related trace functions, with conditions for equality, *Rev. Math. Phys.* **22** (2010) 1099–1121.

[19] A. Jenčová, The structure of strongly additive states and Markov triplets on the CAR algebra, *J. Math. Phys.* **51** (2010) 112103.

[20] A. Jenčová, D. Petz and J. Pitrik, Markov triplets on CCR-algebras, *Acta Sci. Math.* (*Szeged*) **76** (2010) 27–50.

[21] E. H. Lieb and M. B. Ruskai, Some operator inequalities of the Schwarz type, *Adv. Math.* **12** (1974) 269–273.
[22] M. Mosonyi and D. Petz, Structure of sufficient quantum coarse grainings, *Lett. Math. Phys.* **68** (2004) 19–30.
[23] H. Nagaoka, The converse part of the theorem for quantum Hoeffding bound (2006); quant-ph/0611289.
[24] M. Ohya and D. Petz, *Quantum Entropy and Its Use*, 2nd edn. (Springer-Verlag, 2004).
[25] V. Paulsen, *Completely Bounded Maps and Operator Algebras* (Cambridge University Press, 2002).
[26] D. Petz, A dual in von Neumann algebras with weights, *Quart. J. Math. Oxford* **35** (1984) 475–483.
[27] D. Petz, Sufficient subalgebras and the relative entropy of states of a von Neumann algebra, *Comm. Math. Phys.* **105** (1986) 123–131.
[28] D. Petz, Quasi-entropies for finite quantum systems, *Rep. Math. Phys.* **21** (1986) 57–65.
[29] D. Petz, Sufficiency of channels over von Neumann algebras, *Quart. J. Math. Oxford* **39** (1988) 907–1008.
[30] D. Petz, Monotone metrics on matrix spaces, *Linear Algebra Appl.* **244** (1996) 81–96.
[31] M. E. Shirokov, On strict decrease of the Holevo quantity (2011); arXiv:1106.3297.
[32] K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf and F. Verstraete, The $\chi^2$-divergence and mixing times of quantum Markov processes, *J. Math. Phys.* **51** (2010) 122201.
[33] H. Strasser, *Mathematical Theory of Statistics. Statistical Experiments and Asymptotic Decision Theory* (Walter de Gruyter, Berlin, 1985).
[34] M. Takesaki, Conditional expectations in von Neumann algebras, *J. Funct. Anal.* **9** (1972) 306–321.

Vol. 70 (2012) *REPORTS ON MATHEMATICAL PHYSICS* No. 2

# COMPARISON OF QUANTUM BINARY EXPERIMENTS

ANNA JENČOVÁ*

Mathematical Institute, Slovak Academy of Sciences,
Štefánikova 49, 814 73 Bratislava, Slovakia
(e-mail: jenca@mat.savba.sk)

(*Received April 5, 2012 – Revised June 21, 2012*)

A quantum binary experiment consists of a pair of density operators on a finite-dimensional Hilbert space. An experiment $\mathcal{E}$ is called $\epsilon$-deficient with respect to another experiment $\mathcal{F}$ if, up to $\epsilon$, its risk functions are not worse than the risk functions of $\mathcal{F}$, with respect to all statistical decision problems. It is known in the theory of classical statistical experiments that (1) for pairs of probability distributions, one can restrict oneself to testing problems in the definition of deficiency and (2) that 0-deficiency is a necessary and sufficient condition for existence of a stochastic mapping that maps one pair onto another. We show that in the quantum case, the property (1) holds precisely if $\mathcal{E}$ consist of commuting densities. As for property (2), we show that if $\mathcal{E}$ is 0-deficient with respect to $\mathcal{F}$, then there exists a completely positive mapping that maps $\mathcal{E}$ onto $\mathcal{F}$, but it is not necessarily trace preserving.

**Keywords:** comparison of statistical experiments, quantum binary experiments, deficiency, statistical morphisms.

## 1. Introduction

In classical statistics, a statistical experiment is a parametrized family of probability distributions on a sample space $(X, \Sigma)$. The theory of experiments and their comparison was introduced by Blackwell [2] and further developed by many authors, e.g. Torgersen, [17, 18]. Most of the results needed here can be found in [16].

For our purposes, a classical *statistical experiment* $\mathcal{E} = (X, \{p_\theta, \ \theta \in \Theta\})$ is a parametrized set of probability distributions $p_\theta, \theta \in \Theta$, over a finite set $X$, where $\Theta$ is a finite set of parameters. This can be interpreted as follows: $X$ is a set of possible outcomes $x \in X$ of some experiment, each occurring with probability $p(x)$, where $p$ is a member of a parametrized family $\{p_\theta\}$, but the value of the parameter is not known. After observing $x$, a decision $d$ is chosen from a finite set $D$ of possible decisions, with some probability $\mu(x, d)$. The function $\mu : X \times D \to [0, 1]$ is called the *decision function*. It is clear that a decision function is a Markov kernel (or a stochastic matrix), that is, $d \mapsto \mu(x, d)$ is a probability distribution for all $x \in X$.

*Supported by the grants VEGA 2/0032/09 and meta-QUTE ITMS 26240120022.

A *loss function* $W : \Theta \times D \to \mathbb{R}^+$ represents the loss suffered if $d \in D$ is chosen and the true value of the parameter is $\theta$. The *risk*, or the average loss of the decision procedure $\mu$ when the true value is $\theta$ is computed as

$$R_{\mathcal{E}}(\theta, W, \mu) = \sum_{x,d} W_\theta(d)\mu(x, d)p_\theta(x).$$

The couple $(D, W)$ is called a *decision problem*. If $D$ consists of two points, then the decision problems $(D, W)$ are precisely the problems of hypothesis testing.

Let $\mathcal{F}$ be another experiment with the same set of parameters, then its "informative value" can be compared to that of $\mathcal{E}$ by comparing their risk functions for all decision problems. This leads to the definitions of $(k, \epsilon)$-deficiency and $\epsilon$-deficiency, see Section 2. One of the most important results of the theory is the following *randomization criterion*.

THEOREM 1. *Let* $\mathcal{E} = (X, \{p_\theta, \theta \in \Theta\})$ *and* $\mathcal{F} = (Y, \{q_\theta, \theta \in \Theta\})$ *be two experiments. Then* $\mathcal{E}$ *is* $\epsilon$*-deficient with respect to* $\mathcal{F}$ *if and only if there is a Markov kernel* $\lambda : X \times Y \to [0, 1]$ *such that*

$$\|\lambda(p_\theta) - q_\theta\|_1 \leq 2\epsilon,$$

*where* $\lambda(p) = \sum_x \lambda(x, y)p(x)$.

For $\epsilon = 0$, this is the Blackwell–Sherman–Stein theorem [2, 13, 15]. For a general $\epsilon$ it was proved in [17].

If $\Theta$ consists of two points, then the experiment is called *binary*. In this case, $\epsilon$-deficiency is equivalent to $(2, \epsilon)$-deficiency [17], which means that such experiments can be compared by considering only the risk functions of hypothesis testing problems.

The development of the quantum version of comparison of statistical experiments was started recently by several authors [14, 3, 8]. A quantum statistical experiment is a set of density operators on a Hilbert space, mostly of finite dimension. Some versions of the randomization criterion, resp. the Blackwell–Sherman–Stein theorem were obtained, in particular, conditions were found for the existence of a trace preserving completely positive map that maps one experiment onto the other. It was conjectured in [14] that the existence of such positive (but not necessarily completely positive) trace preserving map is equivalent to 0-deficiency. A weaker form of this was obtained in [3], where the notion of a *statistical morphism* was introduced. The (even weaker) notion of a *k-statistical morphism* was considered in [8].

The present paper reviews some of the results of [3] and [8], with focus on the problem of comparison of binary experiments. As an extension of [8], we prove that $(2, \epsilon)$-deficiency and $\epsilon$-deficiency of a quantum experiment $\mathcal{E}$ with respect to another quantum experiment $\mathcal{F}$ are equivalent for any $\mathcal{F}$ precisely if the experiment $\mathcal{E}$ is abelian, that is, all density matrices $\rho_\theta$ commute. Moreover, we use the results in [12] to show that any $k$-statistical morphism can be extended to a map that is completely positive, but not trace preserving in general.

## 2. Quantum statistical experiments

Let $\mathcal{H}$ be a finite-dimensional Hilbert space and let $\mathcal{A} \subseteq B(\mathcal{H})$ be a $C^*$-algebra. Let $\mathcal{S}(\mathcal{A})$ denote the set of density operators in $\mathcal{A}$. A (quantum) statistical experiment $\mathcal{E}$ consists of $\mathcal{A}$ and a family $\{\rho_\theta, \theta \in \Theta\} \subset \mathcal{S}(\mathcal{A})$, which is written as $\mathcal{E} = (\mathcal{A}, \{\rho_\theta, \theta \in \Theta\})$. Throughout the paper, we suppose that $\Theta$ is a finite set.

The family $\{\rho_\theta, \theta \in \Theta\}$ represents our knowledge of the state of the quantum system represented by $\mathcal{A}$: it is known that this family contains the state of the system but the true value of $\theta$ is not known.

Let $(D, W)$ be a decision problem. The decision is made by a measurement on $\mathcal{A}$ with values in $D$. Any such measurement is given by a positive operator-valued measure (POVM) $M : D \to \mathcal{A}$, that is, a collection of operators $M = \{M_d, d \in D\} \subset \mathcal{A}^+$ such that $\sum_d M_d = I$. If all $M_d$ are projections, we say that $M$ is a projection-valued measure (PVM). We will denote the set of all measurements by $\mathcal{M}(D, \mathcal{E})$.

Note that any POVM defines a positive trace preserving map $M : \mathcal{A} \to \mathcal{F}(D)$, where $\mathcal{F}(D)$ is the $C^*$-algebra of all functions $D \to \mathbb{C}$. The map is given by

$$M(a)(d) = \mathrm{Tr}\, M_d a, \qquad a \in \mathcal{A},\ d \in D,$$

and any positive trace preserving map $\mathcal{A} \to \mathcal{F}(D)$ is obtained in this way. Moreover, we define the map $\hat{M} : \mathcal{F}(D) \to \mathcal{A}$ by

$$\hat{M}(f) = \sum_d f(d)(\mathrm{Tr}\, M_d)^{-1} M_d, \qquad f \in \mathcal{F}(D).$$

Then $\hat{M}$ is again positive and trace preserving. Since $\mathcal{F}(D)$ is abelian, both $M$ and $\hat{M}$ are also completely positive [10].

As it was pointed out in [3], the set of quantum experiments contains the set of classical experiments and these correspond precisely to abelian experiments, that is, experiments such that all densities in the family $\{\rho_\theta, \theta \in \Theta\}$ commute. Indeed, let $\mathcal{E}$ be abelian and let $\mathcal{C}$ be the subalgebra generated by $\{\rho_\theta, \theta \in \Theta\}$. Then $\mathcal{C}$ is generated by a PVM $P$ concentrated on a finite set $X$ and we have the classical experiment $(X, \{p_\theta := P(\rho_\theta), \theta \in \Theta\})$. Conversely, let $(Y, \{q_\theta, \theta \in \Theta\})$ be any classical experiment with $|Y| \leq \dim(\mathcal{H})$ and let $Q : Y \to \mathcal{A}$ be any PVM, then $(\mathcal{A}, \{\hat{Q}(q_\theta), \theta \in \Theta\})$ defines an abelian quantum experiment. It is clear that $p_\theta = P(\rho_\theta)$ and $\rho_\theta = \hat{P}(p_\theta)$, $\theta \in \Theta$, so that $\mathcal{E}$ and $(X, \{p_\theta\})$ are mapped onto each other by completely positive trace preserving maps. In particular, the experiments are equivalent in the sense defined below.

## 3. Deficiency

Let $\mathcal{E}$ be an experiment and let $(D, W)$ be a decision problem. The *risk* of the decision procedure $M \in \mathcal{M}(D, \mathcal{E})$ at $\theta$ is computed as [5]

$$R_{\mathcal{E}}(\theta, W, M) = \sum_{d \in D} M(\rho_\theta)(d) W_\theta(d) = \sum_d W_\theta(d) \mathrm{Tr}\, \rho_\theta M_d.$$

Let now $\mathcal{F} = (\mathcal{B}, \{\sigma_\theta, \theta \in \Theta\})$ be another experiment, with $\mathcal{B} \subset B(\mathcal{K})$ for a finite-dimensional Hilbert space $\mathcal{K}$ and with the same parameter set. Let $k \in \mathbb{N}$, $D_k := \{0, \ldots, k-1\}$ and let $\epsilon \geq 0$. We say that $\mathcal{E}$ is *$(k,\epsilon)$-deficient with respect to $\mathcal{F}$*, in notation $\mathcal{E} \geq_{k,\epsilon} \mathcal{F}$, if for every decision problem $(D_k, W)$ (equivalently, for all decision problems $(D, W)$ with $|D| = k$) and every $N \in \mathcal{M}(D_k, \mathcal{F})$, there is some $M \in \mathcal{M}(D_k, \mathcal{E})$ such that

$$R_{\mathcal{E}}(\theta, W, M) \leq R_{\mathcal{F}}(\theta, W, N) + \epsilon \|W_\theta\|, \qquad \theta \in \Theta,$$

where $\|W_\theta\| = \sup_{x \in D_k} W_\theta(x)$. We say that $\mathcal{E}$ is *$\epsilon$-deficient with respect to $\mathcal{F}$*, $\mathcal{E} \geq_\epsilon \mathcal{F}$, if it is $(k,\epsilon)$-deficient for all $k \in \mathbb{N}$.

The relation $\leq_0$ defines a preorder on the set of all experiments. If we have $\mathcal{E} \geq_0 \mathcal{F}$ and simultaneously $\mathcal{F} \geq_0 \mathcal{E}$, then we say that $\mathcal{E}$ and $\mathcal{F}$ are *equivalent*, $\mathcal{E} \sim \mathcal{F}$. The equivalence relation $\mathcal{E} \sim_k \mathcal{F}$ is defined similarly, and $\mathcal{E}$ and $\mathcal{F}$ are called *$k$-equivalent*.

The Theorem 2 below (apart from (iii)) was proved in [8, Theorem 5] in a more general setting. We give the proof in our simpler case, just for the convenience of the reader.

The most important ingredient of the proof is the *minimax theorem*, which can be found in [16].

THEOREM 2. *Let $\mathcal{E} = (\mathcal{A}, \{\rho_\theta, \theta \in \Theta\})$ and $\mathcal{F} = (\mathcal{B}, \{\sigma_\theta, \theta \in \Theta\})$ be two experiments with the same parameter set $\Theta$, $|\Theta| < \infty$. Let $k \in \mathbb{N}$, $\epsilon \geq 0$. The following are equivalent.*

(i) $\mathcal{E} \geq_{k,\epsilon} \mathcal{F}$

(ii) *For every loss function $W : \Theta \times D_k \to \mathbb{R}^+$,*

$$\min_{M \in \mathcal{M}(D_k,\mathcal{E})} \sum_\theta R_{\mathcal{E}}(\theta, W, M) \leq \min_{N \in \mathcal{M}(D_k,\mathcal{F})} \sum_\theta R_{\mathcal{F}}(\theta, W, N) + \epsilon \|W\|$$

*where $\|W\| = \sum_\theta \|W_\theta\|$.*

(iii) *For every loss function $W : \Theta \times D_k \to \mathbb{R}^+$,*

$$\max_{M \in \mathcal{M}(D_k,\mathcal{E})} \sum_\theta R_{\mathcal{E}}(\theta, W, M) \geq \max_{N \in \mathcal{M}(D_k,\mathcal{F})} \sum_\theta R_{\mathcal{F}}(\theta, W, N) - \epsilon \|W\|.$$

(iv) *For every $N \in \mathcal{M}(D_k, \mathcal{F})$ there is some $M \in \mathcal{M}(D_k, \mathcal{E})$ such that*

$$\|M(\rho_\theta) - N(\sigma_\theta)\|_1 \leq 2\epsilon, \qquad \forall \theta \in \Theta.$$

*Proof*: Suppose (i), then for any $N \in \mathcal{M}(D_k, \mathcal{F})$, there is some $M \in \mathcal{M}(D_k, \mathcal{E})$ such that

$$\sum_\theta R_{\mathcal{E}}(\theta, W, M) \leq \sum_\theta R_{\mathcal{F}}(\theta, W, N) + \epsilon \|W\|,$$

this implies (ii).

Suppose (ii) and let $W : \Theta \times D_k \to \mathbb{R}^+$ be a loss function. Then $\tilde{W} : \Theta \times D_k \to \mathbb{R}^+$ given by $\tilde{W}_\theta = \|W_\theta\| - W_\theta$ is a loss function with $\|\tilde{W}\| \leq \|W\|$. Since $R_{\mathcal{E}}(\theta, \tilde{W}, M) = \|W_\theta\| - R_{\mathcal{E}}(\theta, W, M)$ and similarly for $R_{\mathcal{F}}$, we have (ii) implies (iii).

Suppose (iii), and let $N \in \mathcal{M}(D_k, \mathcal{F})$. Then for every loss function $W$, we have

$$\max_{M \in \mathcal{M}(D_k, \mathcal{E})} \sum_\theta R_\mathcal{E}(\theta, W, M) \geq \sum_\theta R_\mathcal{F}(\theta, W, N) - \epsilon \|W\|,$$

and this implies that

$$\sup_{W, \|W\| \leq 1} \min_{M \in \mathcal{M}(D_k, \mathcal{E})} \sum_\theta (R_\mathcal{F}(\theta, W, N) - R_\mathcal{E}(\theta, W, M)) \leq \epsilon.$$

The set $\mathcal{M} = \mathcal{M}(D_k, \mathcal{E})$ is compact and obviously convex and the set $\mathcal{W}$ of all loss functions $W$ with $\|W\| \leq 1$ is convex as well. Moreover, the function $(M, W) \mapsto \sum_\theta (R_\mathcal{F}(\theta, W, N) - R_\mathcal{E}(\theta, W, M))$ is linear in both arguments, hence the minimax theorem applies and we get

$$\begin{aligned} \epsilon &\geq \min_{M \in \mathcal{M}} \sup_{W \in \mathcal{W}} \sum_\theta (R_\mathcal{F}(\theta, W, N) - R_\mathcal{E}(\theta, W, M)) \\ &= \min_{M \in \mathcal{M}} \sup_{W \in \mathcal{W}} \sum_{\theta, d} W_\theta(d)(N(\sigma_\theta)(d) - M(\rho_\theta)(d)). \end{aligned}$$

Let $\mathcal{P}(\Theta)$ be the set of all probability measures on $\Theta$ and let $p \in \mathcal{P}(\Theta)$. For $M \in \mathcal{M}$ fixed, let $W$ be given by

$$W_\theta(x) = \begin{cases} p(\theta) \text{ if } N(\sigma_\theta)(x) - M(\rho_\theta)(x) > 0, \\ 0 \qquad\qquad \text{otherwise.} \end{cases}$$

Then $W \in \mathcal{W}$, so that we get

$$\begin{aligned} \epsilon &\geq \min_{M \in \mathcal{M}} \sum_\theta \sum_{x \in D_k} W_\theta(x)(N(\sigma_\theta)(x) - M(\rho_\theta)(x)) \\ &= \min_{M \in \mathcal{M}} \sum_\theta p(\theta) \frac{1}{2} \|N(\sigma_\theta) - M(\rho_\theta)\|_1. \end{aligned}$$

Since this holds for any $p \in \mathcal{P}(\Theta)$, we have obtained

$$\sup_{p \in \mathcal{P}(\Theta)} \min_{M \in \mathcal{M}} \sum_\theta p(\theta) \|M(\rho_\theta) - N(\sigma_\theta)\|_1 \leq 2\epsilon.$$

The set $\mathcal{P}(\Theta)$ is convex and the function $\mathcal{M} \times \mathcal{P}(\Theta) \to \mathbb{R}$, given by $(M, p) \mapsto \sum_\theta p(\theta) \|M(\rho_\theta) - N(\sigma_\theta)\|_1$ is convex in $M$ and concave (linear) in $p$. Hence the minimax theorem applies again and we have

$$\min_M \sup_p \|M(\rho_\theta) - N(\sigma_\theta)\|_1 = \sup_p \min_M \sum_\theta p(\theta) \|M(\rho_\theta) - N(\sigma_\theta)\|_1 \leq 2\epsilon$$

which clearly implies (iv), by taking the probability measures concentrated in $\theta \in \Theta$.

Suppose (iv) and let $N \in \mathcal{M}(D_k, \mathcal{F})$. Let $M \in \mathcal{M}(D_k, \mathcal{E})$ be chosen for $N$ by (iv). Then for any loss function $W$,

$$R_{\mathcal{E}}(\theta, W, M) - R_{\mathcal{F}}(\theta, W, N) = \sum_{x \in D_k} W_\theta(x)(M(\rho_\theta)(x) - N(\sigma_\theta)(x))$$
$$\leq \frac{\|W_\theta\|}{2}\|M(\rho_\theta) - N(\sigma_\theta)\|_1 \leq \epsilon \|W_\theta\|.$$

so that $\mathcal{E} \geq_{k,\epsilon} \mathcal{F}$. □

The following corollary is a generalization of the classical randomization criterion to the case when the experiment $\mathcal{F}$ is abelian. In the case when $\epsilon = 0$, it was proved in [3].

COROLLARY 1. *Let $\mathcal{E} = (\mathcal{A}, \{\rho_\theta, \theta \in \Theta\})$ and let $\mathcal{F} = (\mathcal{B}, \{\sigma_\theta, \theta \in \Theta\})$ be abelian. Then $\mathcal{E} \geq_\epsilon \mathcal{F}$ if and only if there is a completely positive trace preserving map $T : \mathcal{A} \to \mathcal{B}$ such that*

$$\|T(\rho_\theta) - \sigma_\theta\|_1 \leq 2\epsilon, \quad \theta \in \Theta.$$

*Proof*: Let $(X, \{p_\theta, \theta \in \Theta\})$ be a classical experiment equivalent to $\mathcal{F}$ and let $P = (P_1, \dots, P_m)$ be the PVM such that $P(\sigma_\theta) = p_\theta$, $\theta \in \Theta$. Suppose $\mathcal{E} \geq_\epsilon \mathcal{F}$, then $P \in \mathcal{M}(X, \mathcal{F})$ and by Theorem 2 (iv), there is some $M \in \mathcal{M}(X, \mathcal{E})$ such that

$$\|M(\rho_\theta) - P(\sigma_\theta)\|_1 = \|M(\rho_\theta) - p_\theta\|_1 \leq 2\epsilon.$$

Put $T = \hat{P} \circ M$, then $T : \mathcal{A} \to \mathcal{B}_0 \subseteq \mathcal{B}$ is positive and trace preserving, where $\mathcal{B}_0$ is the abelian subalgebra generated by $P$. Hence $T$ is also completely positive. Moreover,

$$\|T(\rho_\theta) - \sigma_\theta\|_1 = \|\hat{P}(M(\rho_\theta) - p_\theta)\|_1 \leq \|M(\rho_\theta) - p_\theta\|_1 \leq 2\epsilon.$$

For the converse, let $N \in \mathcal{M}(D, \mathcal{F})$ for any finite set $D$. Put $Q = N \circ T$, then $Q \in \mathcal{M}(D, \mathcal{E})$ and

$$\|Q(\rho_\theta) - N(\sigma_\theta)\|_1 = \|N(T(\rho_\theta) - \sigma_\theta)\|_1 \leq 2\epsilon.$$

By Theorem 2 (iv), this implies $\mathcal{E} \geq_\epsilon \mathcal{F}$. □

### 3.1. Deficiency with respect to testing problems

Let $(D_2, W)$ be a decision problem. Then any $M \in \mathcal{M}(D_2, \mathcal{E})$ has the form $(M_0, I - M_0)$ for some $0 \leq M_0 \leq I$ and the risk of $M$ is

$$R_{\mathcal{E}}(\theta, M, W) = W_\theta(1) + (W_\theta(0) - W_\theta(1))\mathrm{Tr}\,\rho_\theta M_0.$$

By Theorem 2 (iii), $\mathcal{E} \geq_{2,\epsilon} \mathcal{F}$ if and only if

$$\max_{\substack{M_0 \in \mathcal{A}, \\ 0 \leq M_0 \leq 1}} \mathrm{Tr} \sum_\theta A_\theta \rho_\theta M_0 \geq \max_{\substack{N_0 \in \mathcal{B}, \\ 0 \leq N_0 \leq 1}} \mathrm{Tr} \sum_\theta A_\theta \sigma_\theta N_0 - \epsilon \|W\| \tag{1}$$

for all loss functions $W$, where we denote $A_\theta := W_\theta(0) - W_\theta(1)$. It is easy to see that

$$\max_{0\le M_0\le 1} \mathrm{Tr} \sum_\theta A_\theta \rho_\theta M_0 = \mathrm{Tr}\Big[\sum_\theta A_\theta \rho_\theta\Big]^+ = \frac{1}{2}\Big(\sum_\theta A_\theta + \Big\|\sum_\theta A_\theta \rho_\theta\Big\|_1\Big), \qquad (2)$$

here we have used the equality $\mathrm{Tr}\, a^+ = \frac{1}{2}(\mathrm{Tr}\, a + \mathrm{Tr}\,|a|)$ for a self-adjoint element $a \in \mathcal{A}$.

THEOREM 3. *$\mathcal{E} \ge_{2,\epsilon} \mathcal{F}$ if and only if*

$$\Big\|\sum_\theta A_\theta \rho_\theta\Big\|_1 \ge \Big\|\sum_\theta A_\theta \sigma_\theta\Big\|_1 - 2\varepsilon \sum_\theta |A_\theta|$$

*for any coefficients $A_\theta \in \mathbb{R}$.*

*Proof*: Follows from (1) and (2). For the 'if' part, put $A_\theta = W_\theta(0) - W_\theta(1)$, we then have $\sum_\theta |A_\theta| \le \|W\|$. For the converse, let $F_+ := \{\theta, A_\theta > 0\}$, $F_- := \{\theta, A_\theta \le 0\}$ and put

$$W_\theta(0) = \begin{cases} A_\theta \text{ if } \theta \in F_+ \\ 0 \ \text{ otherwise} \end{cases},$$

$$W_\theta(1) = \begin{cases} -A_\theta \text{ if } \theta \in F_- \\ 0 \ \text{ otherwise} \end{cases}.$$

Then $W$ is a loss function with $\|W\| = \sum_\theta |A_\theta|$. $\square$

### 3.2. Deficiency and sufficiency

Let $T : \mathcal{A} \to \mathcal{B}$ be a completely positive trace preserving map. The experiment $\mathcal{F} = (\mathcal{B}, \{T(\rho_\theta), \theta \in \Theta\})$ is called a *randomization* of $\mathcal{E}$. If $N \in \mathcal{M}(D, \mathcal{F})$, then $T^*(N) \in \mathcal{M}(D, \mathcal{E})$ and it is clear that $T^*(N)$ has the same risks as $N$, hence $\mathcal{E}$ is 0-deficient with respect to $\mathcal{F}$.

Suppose that in this setting, $\mathcal{F}$ is $(k, 0)$-deficient with respect to $\mathcal{E}$, then we say that $T$ is *$k$-sufficient* for $\mathcal{E}$. If also $\mathcal{E}$ is a randomization of $\mathcal{F}$, then we say that $T$ is *sufficient* for $\mathcal{E}$, this definition of sufficiency was introduced in [11]. If $T$ is a restriction to a subalgebra $\mathcal{A}_0 \subset \mathcal{A}$, then we say that $\mathcal{A}_0$ is $k$-sufficient, resp. sufficient for $\mathcal{E}$, if $T$ is. If the experiments are abelian, then it follows by the randomization criterion that $T$ is sufficient if and only if it is $k$-sufficient for every $k \in \mathbb{N}$. Moreover, for abelian binary experiments, $T$ is sufficient if and only if it is 2-sufficient. (In fact, the last statement hold for all classical statistical experiments [16].)

It is not clear if any of the above two statements holds for quantum experiments. The latter condition for binary experiments was investigated in [6], for a subalgebra $\mathcal{A}_0$. It was shown that $\mathcal{A}_0$ is 2-sufficient if and only if it contains all projections

$P_{t,+}$, $t \geq 0$ (see Lemma 1) and that this is equivalent to sufficiency in some cases. In particular:

THEOREM 4. *Let $\mathcal{E} = (\mathcal{A}, \{\rho_1, \rho_2\})$ be an experiment and let $\mathcal{A}_0 \subseteq \mathcal{A}$ be an abelian subalgebra. Then the following are equivalent:*
(i) *$\mathcal{A}_0$ is 2-sufficient,*
(ii) *$\mathcal{A}_0$ is sufficient,*
(iii) *$\mathcal{A}_0$ is sufficient and $\mathcal{E}$ is abelian.*

*Proof*: The equivalence of (i) and (ii) was proved in [6, Theorem 5(2)], (ii) $\Longrightarrow$ (iii) follows from [9, Theorem 9.10]. (iii) $\Longrightarrow$ (i) is obvious. □

## 4. Binary experiments

Let $\mathcal{E} = (\mathcal{A}, \{\rho_1, \rho_2\})$ be a binary experiment. Note that we may suppose that $\rho_1 + \rho_2$ is invertible, since $\mathcal{E}$ can be replaced by the experiment $(P\mathcal{A}P, \{\rho_1, \rho_2\})$, where $P = \operatorname{supp}(\rho_1 + \rho_2)$ is the support projection of $\rho_1 + \rho_2$.

Let us denote

$$f_{\mathcal{E}}(t) := \max_{\substack{M \in \mathcal{A}, \\ 0 \leq M \leq I}} \operatorname{Tr}(\rho_1 - t\rho_2)M, \qquad t \in \mathbb{R}.$$

Then by (2),

$$f_{\mathcal{E}}(t) = \operatorname{Tr}(\rho_1 - t\rho_2)_+ = \frac{1}{2}(\|\rho_1 - t\rho_2\|_1 + 1 - t). \tag{3}$$

It is easy to see that Theorem 3 for binary experiments has the following form.

THEOREM 5. *Let $\mathcal{E} = \{\mathcal{A}, \{\rho_1, \rho_2\})$ and $\mathcal{F} = (\mathcal{B}, \{\sigma_1, \sigma_2\})$. Then the following are equivalent:*
(i) *$\mathcal{E} \geq_{2,\epsilon} \mathcal{F}$,*
(ii) *$\|\rho_1 - t\rho_2\|_1 \geq \|\sigma_1 - t\sigma_2\|_1 - 2(1+t)\varepsilon$ for all $t \geq 0$,*
(iii) *$f_{\mathcal{E}}(t) \geq f_{\mathcal{F}}(t) - (1+t)\epsilon$ for all $t \geq 0$.*

We will need some properties of the function $f_{\mathcal{E}}$. First, we state the quantum version of the Neyman–Pearson lemma [4, 5]. For this, let us denote $P_{t,+} := \operatorname{supp}(\rho_1 - t\rho_2)_+$ and $P_{t,0} = \ker(\rho_1 - t\rho_2)$ for $t \geq 0$.

LEMMA 1. *We have $f_{\mathcal{E}}(t) = \operatorname{Tr}(\rho_1 - t\rho_2)M$ for some $M \in \mathcal{A}$, $0 \leq M \leq I$, if and only if*

$$M = P_{t,+} + X, \qquad 0 \leq X \leq P_{t,0}.$$

The proof of the following lemma can be found in Appendix.

LEMMA 2.
(i) *$f_{\mathcal{E}}$ is continuous, convex and $f_{\mathcal{E}}(t) \geq \max\{1-t, 0\}$, $t \in \mathbb{R}$.*
(ii) *$f_{\mathcal{E}}$ is nonincreasing in $\mathbb{R}$. Moreover, $f_{\mathcal{E}}$ is analytic in $\mathbb{R}$ except of some points $0 \leq t_1 < \cdots < t_l$, $l \leq \dim(\mathcal{H})$, where $f_{\mathcal{E}}$ is not differentiable. These are exactly the points where $P_{t,0} \neq 0$.*

We will denote by $\mathcal{T}_{\mathcal{E}} := \{t_1, \ldots, t_l\}$ the set of points defined in (ii).

### 4.1. Deficiency and 2-deficiency for binary experiments

For classical binary experiments, it was proved in [17] that $\mathcal{E} \geq_{2,\epsilon} \mathcal{F}$ is equivalent with $\mathcal{E} \geq_{\epsilon} \mathcal{F}$, so that for comparison of such experiments it is enough to consider all testing problems. We prove below that this equivalence remains true if only $\mathcal{E}$ is abelian, and that this property characterizes abelian binary experiments.

We will need the following lemma.

LEMMA 3. *Let $s_1, s_2 \notin \mathcal{T}_{\mathcal{E}}$, $0 < s_1 < s_2$. Then there is a classical experiment $\mathcal{F} = (X = \{1, 2, 3\}, \{p, q\})$, such that $f_{\mathcal{E}}(t) \geq f_{\mathcal{F}}(t)$ for all $t$ and $f_{\mathcal{E}}(s_i) = f_{\mathcal{F}}(s_i)$, $i = 1, 2$.*

*Proof*: Let us define linear functions $g_i(t) := a_i - tb_i$, $i = 0, \ldots, 3$, where $a_0 = b_0 = 1$, $a_3 = b_3 = 0$ and $a_i = f_{\mathcal{E}}(s_i) - s_i f'_{\mathcal{E}}(s_i)$, $b_i = -f'_{\mathcal{E}}(s_i)$, $i = 1, 2$, so that

$$g_i(t) = f_{\mathcal{E}}(s_i) + (t - s_i) f'_{\mathcal{E}}(s_i)$$

is tangent to $f_{\mathcal{E}}$ at $s_i$, $i = 0, 1, 2$, where we put $s_0 = 0$. Since $f_{\mathcal{E}}$ is convex and $f_{\mathcal{E}}(t) \geq \max\{1 - t, 0\}$, $g_i(t) \leq f(t)$, for all $i$ and $t$. Moreover, since $f_{\mathcal{E}}$ is also nonincreasing, we have for any $t < 0$, $-1 = f'_{\mathcal{E}}(t) \leq f'_{\mathcal{E}}(s_1) \leq f'_{\mathcal{E}}(s_2) \leq 0$ so that $b_0 \geq b_1 \geq b_2 \geq b_3$. Convexity and $f_{\mathcal{E}}(0) = 1$ also imply that

$$\begin{aligned} 1 - a_1 &= 1 - f_{\mathcal{E}}(s_1) + s_1 f'_{\mathcal{E}}(s_1) \geq 0, \\ a_1 - a_2 &= f_{\mathcal{E}}(s_1) - f_{\mathcal{E}}(s_2) - f'_{\mathcal{E}}(s_2)(s_1 - s_2) + s_1(b_1 - b_2) \geq 0, \\ a_2 &= f_{\mathcal{E}}(s_2) + s_2 b_2 \geq 0, \end{aligned}$$

so that $a_0 \geq a_1 \geq a_2 \geq a_3$. Put $p_i := a_{i-1} - a_i$, $q_i := b_{i-1} - b_i$, $i = 1, 2, 3$, then $p = (p_1, p_2, p_3)$ and $q = (q_1, q_2, q_3)$ are probability measures. Let $\mathcal{F} := (\{1, 2, 3\}, \{p, q\})$, then

$$f_{\mathcal{F}}(t) = \sum_{i, p_i - tq_i > 0} p_i - tq_i = \sum_{i, g_{i-1}(t) > g_i(t)} g_{i-1}(t) - g_i(t).$$

Let us now define the points $t'_0, \ldots, t'_3$ as follows. Put $t'_0 := 0$ and for $i = 1, 2, 3$, let $t'_i := t'_{i-1}$ if $g_i = g_{i-1}$, otherwise let $t'_i$ be such that $g_i(t) < g_{i-1}(t)$ for $t < t'_i$ and $g_i(t'_i) = g_{i-1}(t'_i)$. Note that $t'_i \geq 0$, since $g_i(0) \leq g_{i-1}(0)$. Moreover, since $g_i(s_i) = f_{\mathcal{E}}(s_i) \geq g_{i-1}(s_i)$, we have $t'_i \leq s_i$ for $i = 0, 1, 2$. In fact, $t'_i < s_i$ for $i = 1, 2$, since $g_{i-1}(s_i) = g_i(s_i) = f_{\mathcal{E}}(s_i)$ implies $f_{\mathcal{E}} = g_i = g_{i-1}$ in some interval containing $s_i$, so that $t'_i = t'_{i-1} \leq s_{i-1} < s_i$. Similarly, for $i = 2, 3$, $g_i(s_{i-1}) \leq f_{\mathcal{E}}(s_{i-1}) = g_{i-1}(s_{i-1})$, so that we either have $t'_i = t'_{i-1}$ or $t'_i > s_{i-1}$. In the case that $g_2(t) > 0$ for all $t$, we put $t'_3 = \infty$. Putting all together, we have $0 = t'_0 \leq t'_1 < s_1 < t'_2 < s_2 < t'_3 \leq \infty$ and

$$\begin{aligned} f_{\mathcal{F}}(t) &= \sum_{j=i}^{3} g_{j-1}(t) - g_j(t) = g_{i-1}(t), \quad t \in \langle t'_{i-1}, t'_i \rangle, \qquad i = 1, 2, 3, \\ f_{\mathcal{F}}(t) &= 0, \quad t \in \langle t'_3, \infty). \end{aligned}$$

It follows that $f_{\mathcal{F}}(t) \leq f_{\mathcal{E}}(t)$ for all $t$ and $f_{\mathcal{F}}(s_i) = f_{\mathcal{E}}(s_i)$, $i = 1, 2$. $\square$

We will now state the main result of this section.

THEOREM 6. *Let $\mathcal{E} = \{\mathcal{A}, \{\rho_1, \rho_2\})$ be a binary experiment. Then the following are equivalent:*

(i) $\mathcal{E} \geq_{2,\epsilon} \mathcal{F} \iff \mathcal{E} \geq_\epsilon \mathcal{F}$ *for any* $\epsilon \geq 0$ *and any abelian binary experiment* $\mathcal{F}$,
(ii) $\mathcal{E} \geq_{2,\epsilon} \mathcal{F} \iff \mathcal{E} \geq_\epsilon \mathcal{F}$ *for any* $\epsilon \geq 0$ *and any binary experiment* $\mathcal{F}$,
(iii) $\mathcal{E} \geq_{2,0} \mathcal{F} \iff \mathcal{E} \geq_0 \mathcal{F}$ *for any abelian binary experiment* $\mathcal{F}$,
(iv) $\mathcal{E}$ *is abelian.*

*Proof*: Suppose (i) and let $\mathcal{F} = (\mathcal{B}, \{\sigma_1, \sigma_2\})$ be any binary experiment such that $\mathcal{E} \geq_{2,\epsilon} \mathcal{F}$. Let $D$ be a finite set and let $N \in \mathcal{M}(D, \mathcal{F})$. Put $p_i := N(\sigma_i)$, $i = 1, 2$, and let $\mathcal{F}_N := (D, \{p_1, p_2\})$. Then by Theorem 5, we have for each $t \geq 0$,

$$\|\rho_1 - t\rho_2\|_1 \geq \|\sigma_1 - t\sigma_2\|_1 - 2(1+t)\epsilon \geq \|p_1 - tp_2\|_1 - 2(1+t)\epsilon.$$

Hence $\mathcal{E} \geq_{2,\epsilon} \mathcal{F}_N$ and (i) implies that $\mathcal{E} \geq_\epsilon \mathcal{F}_N$. By Corollary 1, there is some $M \in \mathcal{M}(D, \mathcal{E})$ such that

$$\|M(\rho_i) - N(\sigma_i)\|_1 = \|M(\rho_i) - p_i\|_1 \leq 2\epsilon, \quad i = 1, 2.$$

By Theorem 2, $\mathcal{E} \geq_\epsilon \mathcal{F}$ and this implies (ii). (ii) trivially implies (iii).

Suppose (iii). Choose any points $s_1, s_2 \notin \mathcal{T}_\mathcal{E}$, $0 < s_1 < s_2$, then by Lemma 3, there is a classical experiment $\mathcal{F} = (\{1, 2, 3\}, \{p_1, p_2\})$ such that $f_\mathcal{E}(t) \geq f_\mathcal{F}(t)$ for $t \geq 0$ and $f_\mathcal{E}(s_i) = f_\mathcal{F}(s_i)$, $i = 1, 2$. By Theorem 5, this implies that $\mathcal{E} \geq_{2,0} \mathcal{F}$ and by (iii), $\mathcal{E} \geq_0 \mathcal{F}$. By Corollary 1, there is a POVM $M : \{1, 2, 3\} \to \mathcal{A}$ such that $p_k = M(\rho_k)$, $k = 1, 2$. For $i = 1, 2$, put $J_i := \{j \in \{1, 2, 3\}, p_1(j) - s_i p_2(j) > 0\}$, then we have

$$\begin{aligned} f_\mathcal{E}(s_i) = f_\mathcal{F}(s_i) &= \sum_{j \in J_i} p_1(j) - s_i p_2(j) \\ &= \sum_{j \in J_i} \mathrm{Tr}\,(\rho_1 M_j) - s_i \mathrm{Tr}\,(\rho_2 M_j) = \mathrm{Tr}\,(\rho_1 - s_i\rho_2) \sum_{j \in J_i} M_j. \end{aligned}$$

Since $s_i \notin \mathcal{T}_\mathcal{E}$, we have $P_{s_i,0} = 0$ and Lemma 1 implies that $\sum_{j \in J_i} M_j = P_{s_i,+}$. Hence the projection $P_{s_i,+}$ is in the range of $M$. Since for all $j \in \{1, 2, 3\}$ we either have $M_j \leq P_{s_i,+}$ or $M_j \leq I - P_{s_i,+}$, $P_{s_i,+}$ must commute with all $M_j$. In particular, $P_{s_1,+}$ and $P_{s_2,+}$ commute.

Since this can be done for any such $s_1$, $s_2$, it follows that all $\{P_{t,+}, t \notin \mathcal{T}_\mathcal{E}\}$ are mutually commuting projections. Since $t \mapsto P_{t,+}$ is right-continuous, it follows that $P_{t_j,+}$ commutes with all $P_{s,+}$ for $s \notin \mathcal{T}_\mathcal{E}$, and by repeating this argument, $P_{t,+}$ are mutually commuting projections for all $t \geq 0$.

Let now $\mathcal{A}_0$ be the subalgebra generated by $\{P_{t,+}, t \geq 0\}$. Then $\mathcal{A}_0$ is an abelian subalgebra which is 2-sufficient for $\mathcal{E}$. Hence $\mathcal{E}$ must be abelian by Theorem 4.

The implication (iv) $\implies$ (i) was proved by Torgersen [17]. □

REMARK 1. If $\dim(\mathcal{H}) = \dim(\mathcal{K}) = 2$, it was proved in [1] that $\mathcal{E} \geq_{2,0} \mathcal{F}$ if and only if $\mathcal{F}$ is a randomization of $\mathcal{E}$. The above proof shows that if $\dim(\mathcal{K}) \geq 3$ this is no longer true unless $\mathcal{E}$ is abelian.

## 5. Statistical morphisms

Let $S_{\mathcal{E}} := \mathrm{span}\{\rho_\theta, \theta \in \Theta\}$. A *k-statistical morphism* [3, 8] is a linear map $L : S_{\mathcal{E}} \to \mathcal{B}$ such that

(i) $L(\rho_\theta) \in \mathcal{S}(\mathcal{B})$ for all $\theta$,

(ii) for each POVM $N : D_k \to \mathcal{B}$ there is some $M \in \mathcal{M}(D_k, \mathcal{E})$ satisfying

$$\mathrm{Tr}\, L(\rho) N_i = \mathrm{Tr}\, \rho M_i, \qquad i \in D_k, \quad \rho \in S_{\mathcal{E}}.$$

The map $L$ is a *statistical morphism* if it is a $k$-statistical morphism for any $k$. It is clear that any positive trace preserving map $L : \mathcal{A} \to \mathcal{B}$ defines a statistical morphism. The proof of the following proposition appears also in [8].

PROPOSITION 1. *$\mathcal{E} \geq_{k,0} \mathcal{F}$ if and only if there is a k-statistical morphism $L : S_{\mathcal{E}} \to \mathcal{B}$ such that $L(\rho_\theta) = \sigma_\theta$.*

*Proof*: Suppose that $\mathcal{E} \geq_{k,0} \mathcal{F}$ for some $k$, then we also have $\mathcal{E} \geq_{2,0} \mathcal{F}$, and by Theorem 3, this implies $\| \sum_\theta A_\theta \rho_\theta \|_1 \geq \| \sum_\theta A_\theta \sigma_\theta \|_1$ for any $A_\theta \in \mathbb{R}$. Put $L : \rho_\theta \mapsto \sigma_\theta$ and extend to $S_{\mathcal{E}}$ by $L(\sum_\theta A_\theta \rho_\theta) = \sum_\theta A_\theta L(\rho_\theta)$, then $\|L(x)\|_1 \leq \|x\|_1$ for $x \in S_{\mathcal{E}}$, so that $L$ is a well-defined linear map on $S_{\mathcal{E}}$. Theorem 2 (iv) now implies that $L$ is a $k$-statistical morphism. The converse is obvious. □

In [14] and [3], a question was raised whether 0-deficiency is equivalent with the existence of a trace preserving positive map that maps one experiment onto another. It is clear that this question is equivalent with the question if any statistical morphism can be extended to a trace preserving positive map. We show below that if $\mathcal{E}$ and $\mathcal{F}$ are binary experiments, then any $k$-statistical morphism such that $L(\rho_i) = \sigma_i$, $i = 1, 2$ can be extended to even a completely positive map, but Theorem 6 implies that such an extension is not trace preserving in general. This shows that the condition that the map preserves trace cannot be omitted.

Let $t_1$ be as in Lemma 2. Note that

$$t_1 = \max\{t \geq 0, f_{\mathcal{E}}(t) = 1 - t\} = \max\{t \geq 0, \rho_1 - t\rho_2 \geq 0\} \tag{4}$$

and $t_1 = 0$ if and only if $\mathrm{supp}\, \rho_2 \nleq \mathrm{supp}\, \rho_1$. Let us denote

$$t_{\max} := \min\{t \geq 0, f_{\mathcal{E}}(t) = 0\} = \min\{t \geq 0, \rho_1 - t\rho_2 \leq 0\}. \tag{5}$$

Then we have either $t_{\max} = t_l$ or $t_{\max} = \infty$, and the latter happens if and only if $\mathrm{supp}\, \rho_1 \nleq \mathrm{supp}\, \rho_2$. We have

$$t_1 \rho_2 \leq \rho_1 \leq t_{\max} \rho_2 \tag{6}$$

and $t_1$, $t_{\max}$ are extremal values for which the inequality occurs. Equivalently,

$$t_{\max}^{-1} \rho_1 \leq \rho_2 \leq t_1^{-1} \rho_1 \tag{7}$$

with $t_{\max}^{-1}$ and $t_1^{-1}$ extremal. We also remark that $t_1 = \sup(\rho_1/\rho_2)$ and $t_{\max} = \inf(\rho_1/\rho_2)$ as defined in [12].

THEOREM 7. *Let $\mathcal{E} = (\mathcal{A}, \{\rho_1, \rho_2\})$, $\mathcal{F} = (\mathcal{B}, \{\sigma_1, \sigma_2\})$ be binary experiments. If $\mathcal{E} \geq_{2,0} \mathcal{F}$, then there is a completely positive map $T : \mathcal{A} \to \mathcal{B}$ such that $T(\rho_i) = \sigma_i$, $i = 1, 2$.*

*Proof*: Let $\mathcal{E} \geq_{0,2} \mathcal{F}$, then there is a 2-statistical morphism $L : S_{\mathcal{E}} \to \mathcal{B}$, $L(\rho_i) = \sigma_i$, $i = 1, 2$. Moreover, $f_{\mathcal{E}}(t) \geq f_{\mathcal{F}}(t)$ for all $t$. Let $t_1'$ and $t_{\max}'$ be as in (4) and (5) for $\mathcal{F}$. Since $f_{\mathcal{F}}(t) \geq \max\{0, 1 - t\}$, we must have $t_1 \leq t_1'$ and $t_{\max}' \leq t_{\max}$. The rest of the proof is the same as the proof of [12, Theorem 21]:

Let $u, v \in S_{\mathcal{E}}$ be positive elements such that $\ker(u) \not\leq \ker(v)$ and $\ker(v) \not\leq \ker(u)$. Then there are some $\varphi, \psi \in \mathcal{H}$ such that $u\varphi = v\psi = 0$, but $u\psi \neq 0$, $v\varphi \neq 0$. Put

$$T(a) = \frac{\langle \psi, a\psi \rangle}{\langle \psi, u\psi \rangle} L(u) + \frac{\langle \varphi, a\varphi \rangle}{\langle \varphi, v\varphi \rangle} L(v), \quad a \in \mathcal{A},$$

then $T$ is a completely positive extension of $L$. We show that such $u$ and $v$ exist.

Suppose $t_{\max} < \infty$ so that $\operatorname{supp} \rho_1 \leq \operatorname{supp} \rho_2$, then $u := t_{\max}\rho_2 - \rho_1$, $v := \rho_1 - t_1\rho_2$. Then $u, v \geq 0$ and the condition on the kernels follows by extremality of $t_1$ and $t_{\max}$. If $t_{\max} = \infty$ but $t_1 > 0$, then we put $u := t_1^{-1}\rho_1 - \rho_2$ and $v := \rho_2$. Finally, if $t_{\max} = \infty$ and $t_1 = 0$, then we put $u := \rho_1$, $v := \rho_2$. □

REMARK 1. One can see that the extension obtained in the above proof cannot be trace preserving unless $\dim \mathcal{H} = 2$ and $\mathcal{E}$ is abelian.

## Appendix: Proof of Lemma 2

The statement (i) follows easily by definition and (3).

Let $\rho(t) := \rho_1 - t\rho_2$. It can be shown ([7, Chap. II]) that the eigenvalues of $\rho(t)$ are analytic functions $t \mapsto \lambda_i(t)$ for all $t \in \mathbb{R}$. It follows that $\rho(t)$ has a constant number $N$ of distinct eigenvalues $\lambda_1(t), \ldots, \lambda_N(t)$, apart from some exceptional points where some of these eigenvalues are equal, and there is a finite number of such points in any finite interval. Moreover, let $P_i(t)$ be the eigenprojection corresponding to $\lambda_i(t)$ for a non-exceptional point $t$, then $t \mapsto P_i(t)$ can be extended to an analytic function for all $t$ such that, if $s$ is an exceptional point, then the projection corresponding to $\lambda_i(s)$ is given by $\sum_{j, \lambda_j(s) = \lambda_i(s)} P_j(s)$. By continuity, $\operatorname{Tr} P_i(t)$ is a constant, we denote it by $m_i$. If $s$ is not an exceptional point, $m_i$ is the multiplicity of $\lambda_i(s)$.

By differentiating the equation $\operatorname{Tr} \rho(s) P_i(s) = m_i \lambda_i(s)$ one obtains

$$\lambda_i'(s) = -\frac{1}{m_i} \operatorname{Tr} \rho_2 P_i(s). \tag{8}$$

It follows that $\lambda_i(s)$ is nonincreasing for all $s$, moreover, $\lambda_i'(s) = 0$ implies that $\rho_2 P_i(s) = 0$, so that $\rho(t) P_i(s) = \rho(s) P_i(s) = \lambda_i(s) P_i(s)$ for all $t$ and $\lambda_i(s)$ is an eigenvalue of $\rho(t)$ for all $t$. Hence $\lambda_i$ is either strictly decreasing or a constant, which must be nonzero, since we assumed that $\rho_1 + \rho_2$ is invertible. It follows that each $\lambda_i$ hits 0 at most once, so that there is only $l \leq N$ points where $\lambda_i(t) = 0$ for some $i$. Let us denote the points by $0 \leq t_1 < \cdots < t_l$, it is clear that these are exactly the points where $P_{t,0} \neq 0$. Let $J_j := \{i, \lambda_i(t_j) > 0\}$, $j = 1, \ldots, l$. Then

$J_j \subset J_{j-1}$ and

$$f_{\mathcal{E}}(t) = \sum_{i \in J_j} m_i \lambda_i(t), \qquad t \in \langle t_{j-1}, t_j \rangle, \quad j = 1, \ldots, l.$$

This implies (ii). □

REFERENCES

[1] P. Alberti and A. Uhlmann: A problem relating to positive linear map on matrix algebras, *Rep. Math. Phys.* **18,**, (1980), 163–176.
[2] D. Blackwell: Comparison of experiments, Proc. 2nd Berkeley Symp. on Math. Stat. and Probab., 93–102 (1951).
[3] F. Buscemi: Comparison of quantum statistical models: equivalent conditions for sufficiency, *Commun. Math. Phys.* **310,**, (2012), 625–647, arXiv:1004.3794.
[4] C. E. Helstrom: *Quantum Detection and Estimation Theory*, Academic Press, New York 1976.
[5] A. Holevo: *Statistical Structure of Quantum Theory*, Springer-Verlag Berlin Heidelberg (2001)
[6] A. Jenčová: Quantum hypothesis testing and sufficient subalgebras, *Lett. Math. Phys.* **93,**, (2010), 15–27.
[7] T. Kato: *Perturbation Theory for Linear Operators*, Springer 1995.
[8] K. Matsumoto: A quantum version of randomization condition, arXiv:1012.2650.
[9] M Ohya and D. Petz, *Quantum Entropy and its Use*, Springer, Heidelberg (1993).
[10] V. Paulsen: *Completely Bounded Maps and Operator Algebras*, Cambridge University Press, 2002.
[11] D. Petz: Sufficiency of channels over von Neumann algebras, *Quart. J. Math. Oxford* **39** (1988), 97–108.
[12] D. Reeb, M. J. Kastoryano and M. M. Wolf: Hilbert's projective metric in quantum information theory, *J. Math. Phys.* **52**, (2011) 082201.
[13] S. Sherman: On a theorem of Hardy, Littlewood, Pólya and Blackwell, *Proc. Nat. Acad. Sciences* **37,**, (1951), 826–831.
[14] E. Shmaya: Comparison of information structures and completely positive maps, *J. Phys. A: Math. and Gen.* **38,**, (2005), i9717–9127.
[15] C. Stein: *Notes on a Seminar on Theoretical Statistics. I. Comparison of Experiments*, University of Chicago 1951.
[16] H. Strasser: *Mathematical Theory of Statistics*, de Gruyter, Berlin, New York 1985.
[17] E. Torgersen: Comparison of statistical experiments when the parameter space is finite, *Z. Wahrscheintlichkeitstheorie verw. geb.* **16,**, (1970), 219–249.
[18] E. Torgersen, *Comparison of Statistical Experiments*, Cambridge University Press 1991.

# Comparison of quantum channels and statistical experiments

Anna Jenčová
Mathematical Institute, Slovak Academy of Sciences
Štefánikova 49, Bratislava, Slovakia
Email: jenca@mat.savba.sk

***Abstract*—For a pair of quantum channels with the same input space, we show that the possibility of approximation of one channel by post-processings of the other channel can be characterized by comparing the success probabilities for the two ensembles obtained as outputs for any ensemble on the input space coupled with an ancilla. This provides an operational interpretation to a natural extension of Le Cam's deficiency to quantum channels. In particular, we obtain a version of the randomization criterion for quantum statistical experiments. The proofs are based on some properties of the diamond norm and its dual, which are of independent interest.**

## I. Introduction

The theory of comparison of statistical experiments started in the work of Blackwell [1], who introduced a natural ordering of experiments in terms of the risks of optimal decision rules. This ordering was extended by Le Cam [2] into a deficiency measure on statistical experiments, expressing how well an experiment $\mathcal{S}$ can be approximated by randomizations of another experiment $\mathcal{T}$. Le Cam's randomization criterion shows that deficiency also gives the maximal loss in the average payoffs of decision procedures, experienced when the experiment $\mathcal{S}$ is replaced by $\mathcal{T}$. For an account on comparison of statistical experiments, see e.g. [3, 4].

An extension of Blackwell's results for quantum experiments was first obtained by Shmaya [5] in the framework of quantum information structures. In [6], a theory of comparison for both classical and quantum experiments is developed in terms of statistical morphisms. In both works, either additional entanglement or composition of the experiment with a complete set of states is required. Quantum versions of Le Cam's randomization criterion were studied in [7, 8]. In particular, Matsumoto in [8] introduced a natural generalization of classical decision problems to quantum ones and proved a quantum randomization criterion in this setting. The main drawback of this approach is the lack of operational interpretation for quantum decision problems.

Comparison of channels can be obtained as an extension of the theory of comparison of experiments. A natural idea is the following: given two channels with the same input space, compare the two experiments emerging as outputs for a single input experiment. If the output experiment of the channel $\Psi$ is always more informative than the output of the channel $\Phi$, we say that $\Psi$ is less noisy than $\Phi$. An ordering of classical channels was first introduced in the work by Shannon [9], where a coding/decoding criterion was applied. Similar orderings were studied in e.g. [10, 11]. For some more recent works see e.g. [12, 13].

This work was supported by the grants VEGA 2/0069/16 and by Science and Technology Assistance Agency under the contract no. APVV-0178-11.

In the quantum setting, it is possible to use a stronger ordering, namely to consider experiments on the input space coupled with an ancilla. As it turns out, for quantum channels, $\Psi$ is less noisy in this stronger sense if and only if $\Phi$ is a post-processing of $\Psi$. In fact, it is enough to compare guessing probabilities for ensembles of states. This remarkable result was first obtained by Chefles in [14], based on [5]. It was extended and refined in [6], in particular it was proved that no entanglement in the input ensemble is needed. Some applications were already found in [13, 15–17].

The aim of the present work is to establish an approximate version of these results, which may be called the randomization criterion for quantum channels. More precisely, we study an extension of Le Cam's deficiency for quantum channels, based on the diamond norm. Such definitions appear naturally in quantum information theory, for example the approximate (anti)degradable channels, [18]. We show that deficiency can be characterized by comparing success probabilities for output ensembles, with respect to the success probability of the input ensemble. These results are then applied to statistical experiments and a quantum randomization criterion is proved in terms of success probabilities.

The diamond norm appears as a distinguishability norm for quantum channels [19]. As it was observed in [20], this norm can be defined using the order structure given by the cone of completely positive maps. We also show that the dual norm on positive elements can be expressed as the optimal success probability for a certain ensemble. These properties provide a convenient framework for proving our results and are of independent interest.

## II. Notations and preliminaries

If not stated otherwise, the full proofs can be found in [21].

Throughout the paper, all Hilbert spaces are finite dimensional. If $\mathcal{H}$ is a Hilbert space, we fix an orthonormal basis $\{|e_i\rangle, i = 1, \ldots, \dim(\mathcal{H})\}$ in $\mathcal{H}$. We will denote the algebra of linear operators on $\mathcal{H}$ by $B(\mathcal{H})$, the set of positive operators by $B(\mathcal{H})^+$ and the real vector space of self-adjoint elements



978-1-5090-1805-5/16/$31.00 ©2016 IEEE

by $B_h(\mathcal{H})$. The set of states, or density operators, on $\mathcal{H}$ will be denoted by $\mathfrak{S}(\mathcal{H}) := \{\sigma \in B(\mathcal{H})^+,\ \mathrm{Tr}\,\sigma = 1\}$.

Let $\mathcal{L}(\mathcal{H}, \mathcal{K})$ denote the real vector space of Hermitian linear maps $B(\mathcal{H}) \to B(\mathcal{K})$. The set $\mathcal{L}(\mathcal{H}, \mathcal{K})^+$ of completely positive maps forms a closed convex cone in $\mathcal{L}(\mathcal{H}, \mathcal{K})$ which is pointed and generating. With this cone, $\mathcal{L}(\mathcal{H}, \mathcal{K})$ becomes an ordered vector space. We will denote the corresponding order by $\leq$. An element of $\mathcal{L}(\mathcal{H}, \mathcal{K})^+$ that preserves trace is usually called a channel. We will denote the set of all channels by $\mathcal{C}(\mathcal{H}, \mathcal{K})$.

For $\phi \in \mathcal{L}(\mathcal{H}, \mathcal{H})$, we define

$$s(\phi) = \sum_{i,j} \langle e_i, \phi(|e_i\rangle\langle e_j|) e_j\rangle.$$

It is easy to see that $s$ defines a linear functional $\mathcal{L}(\mathcal{H}, \mathcal{H}) \to \mathbb{R}$ and for all $\phi \in \mathcal{L}(\mathcal{H}, \mathcal{K})$, $\psi \in \mathcal{L}(\mathcal{K}, \mathcal{H})$, $s(\psi \circ \phi) = s(\phi \circ \psi)$.

We now identify the dual space of $\mathcal{L}(\mathcal{H}, \mathcal{K})$ with $\mathcal{L}(\mathcal{K}, \mathcal{H})$, where duality is given by

$$\langle \psi, \phi\rangle = s(\psi \circ \phi), \qquad \phi \in \mathcal{L}(\mathcal{H}, \mathcal{K}),\ \psi \in \mathcal{L}(\mathcal{K}, \mathcal{H}).$$

Note that the tracelike property of $s$ implies that we have $\langle \phi, \psi\rangle = \langle \psi, \phi\rangle$ and

$$\langle \phi, \xi \circ \psi\rangle = \langle \phi \circ \xi, \psi\rangle = \langle \psi \circ \phi, \xi\rangle.$$

The dual cone of positive functionals satisfies

$$\begin{aligned}(\mathcal{L}(\mathcal{H}, \mathcal{K})^+)^* &:= \{\psi \in \mathcal{L}(\mathcal{K}, \mathcal{H}), \langle \psi, \phi\rangle \geq 0, \forall \phi \in \mathcal{L}(\mathcal{H}, \mathcal{K})^+\} \\ &= \mathcal{L}(\mathcal{K}, \mathcal{H})^+,\end{aligned}$$

so that the cone of completely positive maps is self-dual.

*Remark* 1. Let us denote

$$X_\mathcal{H} := \sum_{i,j} |e_i\rangle\langle e_j| \otimes |e_i\rangle\langle e_j| \in B(\mathcal{H} \otimes \mathcal{H})^+.$$

The Choi representation $C : \phi \mapsto (\phi \otimes id_\mathcal{H})(X_\mathcal{H})$ provides an order isomorphism of $\mathcal{L}(\mathcal{H}, \mathcal{K})$ onto $B_h(\mathcal{K} \otimes \mathcal{H})$ with the cone of positive operators $B(\mathcal{K} \otimes \mathcal{H})^+$. Note also that for any $\phi \in \mathcal{L}(\mathcal{H}, \mathcal{H})$, $s(\phi) = \mathrm{Tr}\, C(\phi) X_\mathcal{H}$, so that

$$\langle \psi, \phi\rangle = s(\psi \circ \phi) = \mathrm{Tr}\,[C(\psi \circ \phi) X_\mathcal{H}] = \mathrm{Tr}\,[C(\phi) C(\psi^*)].$$

It is of course possible to use the Choi representation with this duality, but for our purposes it is mostly more convenient to work with the spaces of mappings.

## III. The diamond norm and its dual

The diamond norm in $\mathcal{L}(\mathcal{H}, \mathcal{K})$ is defined by

$$\|\phi\|_\diamond = \sup_{\rho \in \mathfrak{S}(\mathcal{H} \otimes \mathcal{H})} \|(\phi \otimes id)(\rho)\|_1, \tag{1}$$

where $\|\cdot\|_1$ denotes the trace norm in $B(\mathcal{K} \otimes \mathcal{H})$. It was proved in [20] that this norm is obtained from the set of channels and the order structure in $\mathcal{L}(\mathcal{H}, \mathcal{K})$. Namely, for $\phi \in \mathcal{L}(\mathcal{H}, \mathcal{K})$,

$$\|\phi\|_\diamond = \inf_{\alpha \in \mathcal{C}(\mathcal{H}, \mathcal{K})} \inf\{\lambda > 0, -\lambda\alpha \leq \phi \leq \lambda\alpha\}. \tag{2}$$

It was also shown that the dual norm in $\mathcal{L}(\mathcal{K}, \mathcal{H})$, which we will denote by $\|\cdot\|^\diamond$, is similarly obtained from the set of erasure channels $\{\phi_\sigma : B(\mathcal{K}) \ni A \mapsto \mathrm{Tr}\,[A]\sigma,\ \sigma \in \mathfrak{S}(\mathcal{H})\}$:

$$\|\psi\|^\diamond = \inf_{\sigma \in \mathfrak{S}(\mathcal{H})} \inf\{\lambda > 0, -\lambda\phi_\sigma \leq \psi \leq \lambda\phi_\sigma\}. \tag{3}$$

We list some useful properties of these norms.

**Proposition 1.** (i) *If $\phi \in \mathcal{L}(\mathcal{H}, \mathcal{K})^+$, then*

$$\|\phi\|_\diamond = \sup_{\sigma \in \mathfrak{S}(\mathcal{H})} \mathrm{Tr}\,[\phi(\sigma)], \quad \|\phi\|^\diamond = \sup_{\alpha \in \mathcal{C}(\mathcal{K}, \mathcal{H})} \langle \alpha, \phi\rangle.$$

(ii) *If $\phi, \psi \in \mathcal{C}(\mathcal{H}, \mathcal{K})$, then*

$$\|\phi - \psi\|_\diamond = 2 \sup_{\gamma \geq 0, \|\gamma\|^\diamond \leq 1} \langle \gamma, \phi - \psi\rangle$$

(iii) *If $\chi \in \mathcal{C}(\mathcal{K}, \mathcal{K}')$ and $\xi \in \mathcal{C}(\mathcal{H}', \mathcal{H})$, then the maps $\phi \mapsto \chi \circ \phi$ and $\phi \mapsto \phi \circ \xi$ are contractions with respect to both $\|\cdot\|_\diamond$ and $\|\cdot\|^\diamond$.*

An important property of the dual norm is its relation to success probabilities for ensembles of quantum states. Let $\mathcal{E} = \{\lambda_i, \sigma_i\}_{i=1}^k$ be and ensemble on $\mathcal{H}$, here $\sigma_i \in \mathfrak{S}(\mathcal{H})$ and $\lambda_1, \dots, \lambda_k$ are prior probabilities. In the setting of multiple hypothesis testing, the task is to guess which one is the true state. Any procedure to obtain such a guess can be identified with some POVM $M = \{M_1, \dots, M_k\}$, $M_i \in B(\mathcal{H})^+$, $\sum_i M_i = I$. Here $\mathrm{Tr}\,\sigma_i M_j$ is interpreted as the probability that $\sigma_j$ is chosen while the true state is $\sigma_i$, so that the average success probability for the procedure $M$ is $\sum_i \lambda_i \mathrm{Tr}\, M_i \sigma_i$. One can show that the maximum probability of a successful guess for this ensemble has the form $P_{succ}(\mathcal{E}) = \|\phi_\mathcal{E}\|^\diamond$, where $\phi_\mathcal{E} \in \mathcal{L}(\mathbb{C}^n, \mathcal{K})$ is the map $A \mapsto \sum_i A_{ii} \lambda_i \sigma_i$. More generally, we have

**Proposition 2.** *Let $\gamma \in \mathcal{L}(\mathcal{K}, \mathcal{H})^+$. Then there is an (equiprobable) ensemble $\mathcal{E}_\gamma$ on $\mathcal{H} \otimes \mathcal{K}$ such that*

$$\|\gamma\|^\diamond = \dim(\mathcal{K}) \mathrm{Tr}\,[\gamma(I)] P_{succ}(\mathcal{E}_\gamma).$$

*Moreover, for any $\phi \in \mathcal{L}(\mathcal{H}, \mathcal{H}')^+$, we have*

$$\mathcal{E}_{\phi\circ\gamma} = (\phi \otimes id)(\mathcal{E}_\gamma).$$

## IV. The main result

Let $\Phi \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ and $\Psi \in \mathcal{C}(\mathcal{H}, \mathcal{K}')$. Similarly to Le Cam's deficiency for statistical experiments, we may define the deficiency of $\Phi$ with respect to $\Psi$ by

$$\delta(\Phi, \Psi) = \inf_{\alpha \in \mathcal{C}(\mathcal{K}', \mathcal{K})} \|\Phi - \alpha \circ \Psi\|_\diamond.$$

Since $\mathcal{C}(\mathcal{K}', \mathcal{K})$ is convex and compact, the infimum is attained, in particular, $\delta(\Phi, \Psi) = 0$ if and only if $\Phi = \alpha \circ \Psi$ for some $\alpha \in \mathcal{C}(\mathcal{K}', \mathcal{K})$. In this case, we write $\Phi \preceq \Psi$ We also define Le Cam distance by

$$\Delta(\Phi, \Psi) = \max\{\delta(\Phi, \Psi), \delta(\Psi, \Phi)\}.$$

This defines a preorder on the set of channels with the same input space. The following data processing inequalities for $\delta$ are obvious consequences of their definition and Proposition 1 (iii).

**Proposition 3.** *Let $\Phi_1, \Phi_2, \Phi, \Psi_1, \Psi_2, \Psi$ be channels with the same input space.*

(i) *If $\Phi_1 \preceq \Phi_2$, then $\delta(\Phi_1, \Psi) \leq \delta(\Phi_2, \Psi)$.*
(ii) *If $\Psi_1 \preceq \Psi_2$, then $\delta(\Phi, \Psi_1) \geq \delta(\Phi, \Psi_2)$.*

Let now $\delta(\Phi, \Psi) = 0$. Then for any ensemble $\mathcal{E}$ on the tensor product $\mathcal{H} \otimes \mathcal{H}_0$ with an ancillary Hilbert space $\mathcal{H}_0$, we have

$$P_{succ}((\Phi \otimes id_{\mathcal{H}_0})(\mathcal{E})) \leq P_{succ}((\Psi \otimes id_{\mathcal{H}_0})(\mathcal{E})). \quad (4)$$

The converse was proved in [6, 14]. Our aim is to prove an $\epsilon$-version of this result.

**Theorem 1.** *Let $\Phi \in \mathcal{C}(\mathcal{H}, \mathcal{K})$, $\Psi \in \mathcal{C}(\mathcal{H}, \mathcal{K}')$, $\epsilon \geq 0$. Then $\delta(\Phi, \Psi) \leq \epsilon$ if and only for any finite dimensional Hilbert space $\mathcal{K}_0$ and any ensemble $\mathcal{E}$ on $\mathcal{H} \otimes \mathcal{K}_0$,*

$$P_{succ}((\Phi \otimes id_{\mathcal{K}_0})(\mathcal{E})) \leq P_{succ}((\Psi \otimes id_{\mathcal{K}_0})(\mathcal{E})) + \frac{\epsilon}{2} P_{succ}(\mathcal{E})$$

*Moreover, one can restrict to $\mathcal{K}_0 = \mathcal{K}$ and equiprobable ensembles with $k = \dim(\mathcal{K})^2$ elements.*

*Proof.* Assume that $\alpha \in \mathcal{C}(\mathcal{K}', \mathcal{K})$ is a channel such that

$$\|\Phi - \alpha \circ \Psi\|_\diamond \leq \epsilon.$$

Then for any $\gamma \in \mathcal{L}(\mathcal{K}_0, \mathcal{H})^+$ and $\chi \in \mathcal{C}(\mathcal{K}, \mathcal{K}_0)$, we have by positivity and Proposition 1 (ii), (iii) that

$$\begin{aligned}\langle \gamma, \chi \circ \Phi \rangle &\leq \langle \gamma, \chi \circ \alpha \circ \Psi \rangle + |\langle \gamma, \chi \circ (\Phi - \alpha \circ \Psi) \rangle| \\ &\leq \langle \gamma, \chi \circ \alpha \circ \Psi \rangle + \frac{1}{2} \|\gamma\|^\diamond \|\chi \circ (\alpha \circ \Psi - \Phi)\|_\diamond \\ &\leq \langle \gamma, \chi \circ \alpha \circ \Psi \rangle + \frac{1}{2} \epsilon \|\gamma\|^\diamond\end{aligned}$$

From this, we obtain by Proposition 1 (i) and properties of $s$ that

$$\begin{aligned}\|\Phi \circ \gamma\|^\diamond &= \sup_{\chi \in \mathcal{C}(\mathcal{K}, \mathcal{K}_0)} \langle \chi, \Phi \circ \gamma \rangle = \sup_{\chi \in \mathcal{C}(\mathcal{K}, \mathcal{K}_0)} \langle \gamma, \chi \circ \Phi \rangle \\ &\leq \sup_{\chi \in \mathcal{C}(\mathcal{K}, \mathcal{K}_0)} \langle \gamma, \chi \circ \alpha \circ \Psi \rangle + \frac{1}{2} \epsilon \|\gamma\|^\diamond \\ &= \sup_{\chi \in \mathcal{C}(\mathcal{K}, \mathcal{K}_0)} \langle \chi \circ \alpha, \Psi \circ \gamma \rangle + \frac{1}{2} \epsilon \|\gamma\|^\diamond \\ &\leq \|\Psi \circ \gamma\|^\diamond + \frac{1}{2} \epsilon \|\gamma\|^\diamond.\end{aligned}$$

Hence we have proved that $\delta(\Phi, \Psi) \leq \epsilon$ implies

$$\|\Phi \circ \gamma\|^\diamond \leq \|\Psi \circ \gamma\|^\diamond + \frac{1}{2} \epsilon \|\gamma\|^\diamond, \quad \forall \gamma \in \mathcal{L}(\mathcal{K}_0, \mathcal{H})^+. \quad (5)$$

Since by (1) we have $\|\phi\|_\diamond = \|\phi \otimes id_{\mathcal{K}_0}\|_\diamond$ for any $\mathcal{K}_0$, we also have $\delta(\Phi \otimes id_{\mathcal{K}_0}, \Psi \otimes id_{\mathcal{K}_0}) \leq \epsilon$. Hence we obtain

$$\|(\Phi \otimes id_{\mathcal{K}_0}) \circ \gamma\|^\diamond \leq \|(\Psi \otimes id_{\mathcal{K}_0}) \circ \gamma\|^\diamond + \frac{\epsilon}{2} \|\gamma\|^\diamond \quad (6)$$

for all $\gamma \in \mathcal{L}(\mathcal{K}_1, \mathcal{H} \otimes \mathcal{K}_0)$ and any $\mathcal{K}_1$. If $\mathcal{E}$ is any ensemble on $\mathcal{H} \otimes \mathcal{K}_0$, then

$$P_{succ}((\Phi \otimes id)(\mathcal{E})) = \|\phi_{(\Phi \otimes id)(\mathcal{E})}\|^\diamond = \|(\Phi \otimes id) \circ \phi_{\mathcal{E}}\|^\diamond$$

and similarly for $\Psi$. Putting $\gamma = \phi_{\mathcal{E}}$ in (6) implies the desired inequality.

For the converse, note that by Proposition 1 (ii), we have

$$\delta(\Phi, \Psi) = 2 \min_{\alpha \in \mathcal{C}(\mathcal{K}', \mathcal{K})} \left\{ \max_{\substack{\gamma \in \mathcal{L}(\mathcal{K}, \mathcal{H})^+, \\ \|\gamma\|^\diamond \leq 1}} \langle \gamma, \Phi - \alpha \circ \Psi \rangle \right\}$$

Since the sets $\mathcal{C}(\mathcal{K}', \mathcal{K})$ and $\{\gamma \in \mathcal{L}(\mathcal{K}, \mathcal{H})^+, \|\gamma\|^\diamond \leq 1\}$ are both convex and compact and the map $(\alpha, \gamma) \mapsto \langle \gamma, \Phi - \alpha \circ \Psi \rangle$ is linear in both variables, we may apply the minimax theorem, see e.g. [3]. It follows that

$$\begin{aligned}\delta(\Phi, \Psi) &= 2 \max_\gamma \min_\alpha \langle \gamma, \Phi - \alpha \circ \Psi \rangle \\ &= 2 \max_\gamma \{\langle \gamma, \Phi \rangle - \|\Psi \circ \gamma\|^\diamond\} \\ &\leq 2 \max_\gamma \{\|\Phi \circ \gamma\|^\diamond - \|\Psi \circ \gamma\|^\diamond\}.\end{aligned}$$

Proposition 2 and the assumption now imply that the last expression is less that $\epsilon$.

□

In the case $\epsilon = 0$, we obtain a stronger condition. Similar results were proved in [6].

**Theorem 2.** *Let $\Phi \in \mathcal{C}(\mathcal{H}, \mathcal{K})$, $\Psi \in \mathcal{C}(\mathcal{H}, \mathcal{K}')$ and let $\xi \in \mathcal{C}(\mathcal{K}_0, \mathcal{K})$ be a surjective channel. Then $\delta(\Phi, \Psi) = 0$ if and only if for any ensemble $\mathcal{E}$ on $\mathcal{H} \otimes \mathcal{K}_0$,*

$$P_{succ}((\Phi \otimes \xi)(\mathcal{E})) \leq P_{succ}((\Psi \otimes \xi)(\mathcal{E})).$$

In particular, by choosing $\xi$ as a classical-to-quantum channel of the form $A \mapsto \sum_i A_{ii} \sigma_i$ for a set of states $\{\sigma_i\}$ that spans $B(\mathcal{K})$, we see that for $\epsilon = 0$ we may restrict to ensembles of separable states.

## V. THE RANDOMIZATION CRITERION FOR QUANTUM EXPERIMENTS

A quantum statistical experiment is a pair $\mathcal{T} = (\mathcal{H}, \{\rho_\theta, \theta \in \Theta\})$, where $\rho_\theta \in \mathfrak{S}(\mathcal{H})$ for all $\theta \in \Theta$ and $\Theta$ is an arbitrary set of parameters. Any experiment can be viewed as the set of possible states of some physical system, determined by some prior information on the true state. Note that this definition contains also classical statistical experiments on finite sample spaces, which can be identified with diagonal density matrices.

Based on the outcome of a measurement on the system, a decision $j$ is chosen from a (finite) set $D$ of decisions. This procedure, or a decision rule, is represented by a POVM $\{M_j, j \in D\}$ on $\mathcal{H}$. The performance of a decision rule is assessed by a payoff function, which in our case is a map $g : \Theta \times D \to \mathbb{R}^+$, representing the payoff obtained if $j \in D$ is chosen while the true state is $\rho_\theta$. The average payoff of the decision rule $M$ at $\theta \in \Theta$ is computed as

$$P_{\mathcal{T}}(\theta, M, g) = \sum_{j \in D} g_{\theta,j} \operatorname{Tr} \rho_\theta M_j.$$

The next theorem is the celebrated Le Cam's randomization criterion for classical statistical experiments. Note that our setting contains only experiments on finite sample spaces, but the theorem holds in a much more general case.

**Theorem 3.** *[2] Let $\mathcal{T}$ and $\mathcal{S}$ be classical statistical experiments. Then the following are equivalent.*

(i) *Tor any decision space $(D, g)$ and any decision rule $M$ for $\mathcal{S}$, there is some decision rule $N$ for $\mathcal{T}$ such that*

$$\sup_{\theta\in\Theta}\left[P_{\mathcal{S}}(\theta, M, g) - P_{\mathcal{T}}(\theta, N, g) - \epsilon \max_d |g(\theta, d)|\right] \leq 0.$$

(ii) *There is some channel $\alpha$ such that*

$$\sup_\theta \|\sigma_\theta - \alpha(\rho_\theta)\|_1 \leq 2\epsilon$$

*Remark* 2. One can show that the condition (i) of the above theorem is equivalent to

$$P_{succ}(\{\lambda_j, \sum_{\theta\in\Theta_0}\mu_\theta^j\sigma_\theta\}) \leq P_{succ}(\{\lambda_j, \sum_{\theta\in\Theta_0}\mu_\theta^j\rho_\theta\}) + \epsilon P_{succ}(\mathcal{E})$$

for any ensemble of the form $\mathcal{E} = \{\lambda_j, \sum_{\theta\in\Theta_0}\mu_\theta^j|e_\theta\rangle\langle e_\theta|\}$ and any finite subset $\Theta_0 \subseteq \Theta$.

As it was proved in [22], Theorem 3 does not hold for quantum experiments. The quantum randomization criterion proved in [8] is based on an extension of the classical decision spaces to quantum ones, but an operational interpretation of the quantum decision problems is not clear. The aim of the present section is to apply Theorem 1 to prove a quantum randomization criterion, formulated in terms of optimal guessing probabilities of some ensembles. In view of Remark 2, this gives a quantum extension of Le Cam's theorem. In the case $\epsilon = 0$, a similar result was obtained in [6].

**Theorem 4.** *Let $\mathcal{S} = (\mathcal{K}, \{\sigma_\theta,\ \theta \in \Theta\})$ and $\mathcal{T} = (\mathcal{H}, \{\rho_\theta,\ \theta\in\Theta\})$ be quantum statistical experiments and let $\epsilon \geq 0$. Then the following are equivalent.*

(i) *There is some $\alpha \in \mathcal{C}(\mathcal{H}, \mathcal{K})$ such that*

$$\sup_{\theta\in\Theta} \|\sigma_\theta - \alpha(\rho_\theta)\|_1 \leq 2\epsilon$$

(ii) *Let $\{\theta_1, \ldots, \theta_n\}$ be any finite subset of $\Theta$ and let $\mathcal{E} = \{\lambda_i, \tau_i\}_{i=1}^k$ be any ensemble on $\mathbb{C}^n \otimes \mathcal{K}$, consisting of block-diagonal states $\tau_i = \sum_{j=1}^n |e_j^n\rangle\langle e_j^n| \otimes \tau_i^j$, $\tau_i^j \in B(\mathcal{K})^+$, $\sum_j \operatorname{Tr}\tau_i^j = 1$. Then*

$$P_{succ}(\{\lambda_i, \sum_{j=1}^n \sigma_{\theta_j}\otimes\tau_i^j\}) \leq$$

$$\leq P_{succ}(\{\lambda_i, \sum_{j=1}^n \rho_{\theta_j}\otimes\tau_i^j\}) + \epsilon P_{succ}(\mathcal{E})$$

*Moreover, in (ii) we may restrict to equiprobable ensembles with $k = \dim(\mathcal{K})^2$.*

*Proof.* Let $\{\theta_1, \ldots, \theta_n\} \subseteq \Theta$ and let $\phi_{\mathcal{S}} \in \mathcal{L}(\mathbb{C}^n, \mathcal{K})$ be given by $A \mapsto \sum_{j=1}^n A_{jj}\sigma_{\theta_j}$. It is easy to see that $\phi_{\mathcal{S}}$ is a channel. Moreover, by [21, Lemma 2], we have for any $\alpha \in \mathcal{C}(\mathcal{H}, \mathcal{K})$,

$$\|\phi_{\mathcal{S}} - \alpha\circ\phi_{\mathcal{T}}\|_\diamond = \max_j \|\sigma_{\theta_j} - \alpha(\rho_{\theta_j})\|_1,$$

where $\phi_{\mathcal{T}}$ is defined analogically. By Theorem 1, the restriction of (i) to $\{\theta_1, \ldots, \theta_n\}$ is equivalent to

$$P_{succ}((\phi_{\mathcal{S}}\otimes id)(\mathcal{E})) \leq P_{succ}((\phi_{\mathcal{T}}\otimes id)(\mathcal{E})) + \epsilon P_{succ}(\mathcal{E})$$

for any ensemble $\mathcal{E}$ on $\mathbb{C}^n\otimes\mathcal{K}$. It is now clear that (i) implies (ii). Since for any state $\rho \in \mathfrak{S}(\mathbb{C}^n\otimes\mathcal{K})$,

$$(\phi_{\mathcal{S}}\otimes id)(\rho) = \sum_j \sigma_{\theta_j}\otimes\tau^j = (\phi_{\mathcal{S}}\otimes id)(\tau)$$

$$(\phi_{\mathcal{T}}\otimes id)(\rho) = \sum_j \rho_{\theta_j}\otimes\tau^j = (\phi_{\mathcal{T}}\otimes id)(\tau)$$

where $\tau = \sum_j |e_j^n\rangle\langle e_j^n|\otimes\tau^j$ is a block diagonal state, we see that (ii) implies that

$$\inf_{\alpha\in\mathcal{C}(\mathcal{H},\mathcal{K})}\sup_{\theta\in\Theta_0}\|\sigma_\theta - \alpha(\rho_\theta)\|_1 \leq 2\epsilon$$

for any finite subset $\Theta_0 \subseteq \Theta$. Let $\mathcal{P}_\Theta$ denote the set of probability measures over $\Theta$ with finite support, then we clearly have

$$\sup_{p\in\mathcal{P}_\Theta}\min_{\alpha\in\mathcal{C}(\mathcal{H},\mathcal{K})}\sum_{\theta\in\Theta}p(\theta)\|\sigma_\theta - \phi(\rho_\theta)\|_1 \leq 2\epsilon.$$

Now we use the minimax theorem once more. For this, note that $\mathcal{P}_\Theta$ is a convex set, $\mathcal{C}(\mathcal{H},\mathcal{K})$ is compact and convex, the map $(p, \alpha) \mapsto \sum_{\theta\in\Theta}p(\theta)\|\sigma_\theta - \alpha(\rho_\theta)\|_1$ is linear in $p$ and continuous and convex in $\alpha$. The minimax theorem can be applied and we obtain

$$\begin{aligned}\sup_{p\in\mathcal{P}_\Theta}\min_{\alpha\in\mathcal{C}(\mathcal{H},\mathcal{K})}&\sum_{\theta\in\Theta}p(\theta)\|\sigma_\theta - \alpha(\rho_\theta)\|_1\\ &= \min_{\alpha\in\mathcal{C}(\mathcal{H},\mathcal{K})}\sup_{p\in\mathcal{P}_\Theta}\sum_{\theta\in\Theta}p(\theta)\|\sigma_\theta - \alpha(\rho_\theta)\|_1\\ &= \min_{\alpha\in\mathcal{C}(\mathcal{H},\mathcal{K})}\sup_{\theta\in\Theta}\|\sigma_\theta - \alpha(\rho_\theta)\|_1.\end{aligned}$$

Hence (ii) implies (i).

□

We will say that an experiment $\mathcal{S}_0 = (\mathcal{K}, \{\tau_\theta, \theta\in\Theta_0\})$ is complete if the set $\{\tau_\theta, \theta\in\Theta_0\}$ spans $B(\mathcal{H})$. If $\Theta_0$ is a finite set, then $\phi_{\mathcal{S}_0}$ is a surjective channel in $\mathcal{C}(\mathbb{C}^{|\Theta_0|}, \mathcal{K})$. By an application of Theorem 2 we obtain

**Corollary 1.** *Let $\mathcal{S} = (\mathcal{K}, \{\sigma_\theta,\ \theta\in\Theta\})$ and $\mathcal{T} = (\mathcal{H}, \{\rho_\theta,\ \theta\in\Theta\})$ be quantum statistical experiments. Let $\mathcal{S}_0 = (\mathcal{K}, \{\tau_1, \ldots, \tau_N\})$ be a complete experiment. Then $\sigma_\theta = \alpha(\rho_\theta)$ for some $\alpha(\mathcal{H},\mathcal{K})$ if and only if for any $\{\theta_1, \ldots, \theta_n\} \subseteq \Theta$ and any collection $\{\Lambda_{j,l}^i\}$, $i = 1, \ldots, k$, $j = 1, \ldots, N$, $l = 1, \ldots, n$ of nonnegative numbers such that $\sum_{j,l}\Lambda_{j,l}^i = 1$ for all $i$, we have*

$$P_{succ}(\{1/k, \sum_{j,l}\Lambda_{j,l}^i\sigma_{\theta_l}\otimes\tau_j\}) \leq P_{succ}(\{1/k, \sum_{j,l}\Lambda_{j,l}^i\rho_{\theta_l}\otimes\tau_j\}).$$

**Corollary 2.** *Let $\Phi \in \mathcal{C}(\mathcal{H},\mathcal{K})$, $\Psi \in \mathcal{C}(\mathcal{H},\mathcal{K}')$ and let $\mathcal{T}_0 = (\mathcal{H}, \{\tau_1^{\mathcal{H}}, \ldots, \tau_M^{\mathcal{H}}\})$, $\mathcal{S}_0 = (\mathcal{K}, \{\tau_1^{\mathcal{K}}, \ldots, \tau_N^{\mathcal{K}}\})$ be complete experiments. Then $\delta(\Phi, \Psi) = 0$ if and only if*

$$P_{succ}((\Phi\otimes id_{\mathcal{K}_0})(\mathcal{E})) \leq P_{succ}((\Psi\otimes id_{\mathcal{K}_0})(\mathcal{E}))$$

*holds for all ensembles of states of the form*

$$\mathcal{E} = \{\lambda_i, \sum_{j,l}\Lambda_{j,l}^i\tau_l^{\mathcal{H}}\otimes\tau_j^{\mathcal{K}}\}.$$

## VI. Concluding remarks

We proved a version of the randomization criterion for quantum channels and applied it to obtain a randomization criterion for quantum statistical experiments. The deficiency $\delta(\Phi, \Psi)$ in some special cases already appeared in quantum information theory and our results can be further used to obtain an operational definition e.g. for the approximately (anti)degradable channels [18], similarly as it was done for antidegradable channels in [15]. Another possible application is to $\epsilon$-private and $\epsilon$-correctable channels [23].

We used some properties of the diamond norm and its dual that can be obtained solely from the order structure given by completely positive maps and the trace preserving condition. This suggests the possibility to apply similar methods to more general situations. For example, one may assume some structure in the channels, obtaining similar results for more specific quantum protocols, such as quantum combs, [24]. It is also possible to define deficiency in terms of pre-processings. In the special case of POVMs regarded as a special kind of channels, this leads to an approximate version of the ordering of POVMs by cleanness, [25]. More generally, the processing can consist of a combination of pre- and post-processing, also allowing some correlations between input and output systems, either classical or quantum. This would be closer to the original definition by Shannon, [9]. It seems that all these situations can be treated within the suggested framework. Another challenging problem is the extension of these results to infinite dimensional Hilbert spaces. Some partial results in this direction were obtained in [26]. Although the methods used in [20] rely on finite dimensions, it seems plausible that the useful properties of the norms can be extended also to this case. All these problems are left for future work.

## References


[1] D. Blackwell, "Comparison of experiments," *Proc. 2nd Berkeley Symp. on Math. Stat. and Probab.*, pp. 93–102, 1951.

[2] L. L. Cam, "Sufficiency and approximate sufficiency," *Ann. Math. Statist.*, vol. 34, pp. 1419–1455, 1964.

[3] H. Strasser, *Mathematical Theory of Statistics*. Berlin, New York: de Gruyter, 1985.

[4] E. Torgersen, *Comparison of Statistical Experiments*. Cambridge University Press, 1991.

[5] E. Shmaya, "Comparison of information structures and completely positive maps," *J. Phys. A-Math. Gen.*, vol. 38, pp. 9717–9727, 2005.

[6] F. Buscemi, "Comparison of Quantum Statistical Models: Equivalent Conditions for Sufficiency," *Commun. Math. Phys.*, vol. 310, pp. 625–647, 2012.

[7] A. Jenčová, "Comparison of quantum binary experiments," *Rep. Math. Phys.*, vol. 70, pp. 237–249, 2012.

[8] K. Matsumoto, "A quantum version of randomization condition," arXiv:1012.2650, 2010.

[9] C. E. Shannon, "A note on a partial ordering for communication channels," *Inform. Control*, vol. 1, pp. 390–397, 1958.

[10] J. Körner and K. Marton, "Comparison of two noisy channels," in *Colloquia Mathematica Societatis Janos Bolyai, Topics in Information Theory*. North Holland, 1977, pp. 411–424.

[11] I. Cziszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Channels*. New York - San Francisco - London: Academic Press, 1981.

[12] M. Raginsky, "Shannon meets Blackwell and Le Cam: Channels, codes, and statistical experiments," *IEEE ISIT Proceedings*, pp. 1220–1224, 2011.

[13] F. Buscemi, "Degradable channels, less noisy channels, and quantum statistical morphisms: an equivalence relation," arXiv:1511.08893, 2015.

[14] A. Chefles, "The quantum Blackwell theorem and minimum error state discrimination," arXiv:0907.0866, 2009.

[15] F. Buscemi, N. Datta, and S. Strelchuk, "Game-theoretic characterization of antidegradable channels," *J. Math. Phys.*, vol. 55, p. 092202, 2014.

[16] F. Buscemi, "Fully quantum second-lawlike statements from the theory of statistical comparisons," arXiv:1505.00535, 2014.

[17] F. Buscemi and N. Datta, "Equivalence between divisibility and monotonic decrease of information in classical and quantum stochastic processes," *Phys. Rev. A*, vol. 93, p. 012101, 2016.

[18] D. Sutter, V. B. Scholz, and R. Renner, "Approximate degradable quantum channels," *IEEE ISIT Proceedings*, pp. 2767–2771, 2015.

[19] A. Kitaev, "Quantum computations: Algorithms and error correction," *Russian Mathematical Surveys*, vol. 52, pp. 1191–1249, 1997.

[20] A. Jenčová, "Base norms and discrimination of generalized quantum channels," *J. Math. Phys*, vol. 55, p. 022201, 2014.

[21] ——, "Comparison of quantum channels and statistical experiments," arXiv:1512.07016, 2015.

[22] K. Matsumoto, "An example of a quantum statistical model which cannot be mapped to a less informative one by any trace preserving positive map," arXiv:1409.5658, 2014.

[23] D. Kretschmann, D. W. Kribs, and R. W. Spekkens, "Complementarity of private and correctable subsystems in quantum cryptography and error correction," *Phys. Rev. A*, vol. 78, p. 032330, 2008.

[24] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Theoretical framework for quantum networks," *Phys. Rev. A*, vol. 80, 2009.

[25] F. Buscemi, M. Keyl, G. M. D'Ariano, P. Perinotti, and R. F. Werner, "Clean positive operator valued measures," *J. Math. Phys.*, vol. 46, 2005.

[26] K. Kaniowski, K. Lubnauer, and A. Luczak, "Quantum Blackwell-Sherman-Stein Theorem and Related Results," *Open Systems & Information Dynamics*, vol. 20, p. 1350017, 2013.

# Chapter 7

# Generalized quantum channels and measurements

[CS1] A. Jenčová, Generalized channels: Channels for convex subsets of the state space, *J. Math. Phys.* **53** (2012), 012201

[CS2] A. Jenčová, Base norms and discrimination of generalized quantum channels, *J. Math. Phys.* **55** (2013), 022201

JOURNAL OF MATHEMATICAL PHYSICS **53**, 012201 (2012)

# Generalized channels: Channels for convex subsets of the state space

Anna Jenčová[a)]
*Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, 814 73 Bratislava, Slovakia*

(Received 26 October 2011; accepted 21 December 2011; published online 25 January 2012)

Let $K$ be a convex subset of the state space of a finite-dimensional $C^*$-algebra. We study the properties of channels on $K$, which are defined as affine maps from $K$ into the state space of another algebra, extending to completely positive maps on the subspace generated by $K$. We show that each such map is the restriction of a completely positive map on the whole algebra, called a generalized channel. We characterize the set of generalized channels and also the equivalence classes of generalized channels having the same value on $K$. Moreover, if $K$ contains the tracial state, the set of generalized channels forms again a convex subset of a multipartite state space, this leads to a definition of a generalized supermap, which is a generalized channel with respect to this subset. We prove a decomposition theorem for generalized supermaps and describe the equivalence classes. The set of generalized supermaps having the same value on equivalent generalized channels is also characterized. Special cases include quantum combs and process positive operator valued measures (POVMs). © *2012 American Institute of Physics*. [doi:10.1063/1.3676294]

## I. INTRODUCTION

The first motivation for this paper comes from the problem of measurement of a quantum channel. A mathematical framework for such measurements, or more generally, for measurements on quantum networks, was introduced in Ref. 4, in terms of testers.[5] For quantum channels, these were called process POVMs or PPOVMs in Ref. 16. Similar to POVMs, a PPOVM is a collection of positive operators $(F_1, \ldots, F_m)$ in the tensor product of the input and output spaces, but summing up to an operator $I_{\mathcal{H}_1} \otimes \omega$ for some state $\omega$ on the input space. The output probabilities of the corresponding channel measurement with values in $\{1, \ldots, m\}$ are then given by

$$p_i(\mathcal{E}) = \mathrm{Tr}\,(M_i X_{\mathcal{E}}), \qquad i = 1, \ldots, m,$$

where $X_{\mathcal{E}}$ is the Choi matrix of the channel $\mathcal{E}$. Via the Choi isomorphism, the set of channels $\mathcal{C}(\mathcal{H}_0, \mathcal{H}_1)$ can be viewed as (a multiple of) an intersection of the set of states in $B(\mathcal{H}_1 \otimes \mathcal{H}_0)$ with a self-adjoint vector subspace $J$. This is a convex set, and a measurement on channels can be naturally defined as an affine map from this set to the set of probability measures on the set of outcomes.

A natural question arising in this context is the following: Are all such affine maps given by PPOVMs? And if so, is this correspondence one-to-one?

Further, the concept of a quantum supermap was introduced in Ref. 6, which is a map $B(\mathcal{H}_1 \otimes \mathcal{H}_0) \to B(\mathcal{H}_3 \otimes \mathcal{H}_2)$ sending channels to channels. It was argued that such a map should be linear and completely positive. But it is clear that it is enough to consider completely positive maps $J \to B(\mathcal{H}_3 \otimes \mathcal{H}_2)$ sending channels to channels. We may then ask whether all such maps extend to a completely positive map on $B(\mathcal{H}_1 \otimes \mathcal{H}_0)$, and if this extension is unique.

Supermaps on supermaps were defined similarly, these are the so-called quantum combs, which are used in description of quantum networks.[4,8] It was proved that all quantum combs can be

[a)]Electronic mail: jenca@mat.savba.sk.

 

represented by memory channels, which are given by a sequence of channels connected by an ancilla, these form the "teeth" of the comb. The theory of quantum combs was subsequently used for optimal cloning[7] and learning[2] of unitary transformations and measurements.[3] As it turns out, the set of all $N$-combs forms again (a multiple of) an intersection of the set of multipartite states by a vector subspace.

To deal with these questions in full generality, we introduce the notion of a channel on a convex subset $K$ of the state space, which is an affine map from $K$ into another state space, extending to a completely positive map on the vector subspace generated by $K$. In order to include all channels, POVMs and instruments, and other similar objects, we work with finite-dimensional $C^*$-algebras rather than matrix algebras. We show that each such map can be extended to a completely positive map on the whole algebra, these maps are called generalized channels (with respect to $K$). Further, a measurement on $K$ is defined as an affine map from $K$ into the set of probability distributions and it is shown that each such measurement is given by a (completely) positive map on the whole algebra if and only if $K$ is a section of the state space, that is, an intersection of the set of states by a linear subspace. This special kind of a generalized channel is called a generalized POVM.

We describe the equivalence class of generalized channels restricting to the same channel on $K$. Moreover, we show that if $K$ contains the tracial state, the set of generalized channels, via Choi representation, is again (a multiple of) a section of some state space, so that we may apply our results on the set of generalized channels themselves and repeat the process infinitely. This leads to the definition of a generalized supermap. We show that the quantum combs and testers are particular cases of generalized supermaps, other examples treated here include channels and measurements on POVMs and PPOVMs, and supermaps on instruments. We also describe channels on the set of states having the same output probabilities for a POVM or a finite number of POVMs.

The outline of the paper is as follows: After Sec. II, we consider extensions of completely positive maps on subspaces of the algebra and of positive affine functions on $K$. If the subspace is self-adjoint and generated by its positive elements, then a consequence of Arveson's extension theorem shows that any completely positive map can be extended to the whole algebra. For positive functionals on $K$, we show that these extend to positive linear functionals on the whole algebra if and only if $K$ is a section of the state space. These results are used in Sect. IV for extension theorems for channels and measurements on $K$. We characterize the generalized channels with respect to $K$ and their equivalence classes. We show that a generalized channel can be decomposed to a so-called simple generalized channel and a channel.

In Sec. V, we prove that the set of generalized channels is again a section of a state space and introduce the generalized supermaps. We give a characterization of generalized supermaps as sections of a multipartite state space and show that the quantum combs are a particular case. We prove a decomposition theorem for the generalized supermaps, similar to the realization of quantum combs by memory channels proved in Ref. 8. In particular, we show that a generalized comb can be decomposed as a simple generalized channel and a comb. Finally, we describe the equivalence classes for generalized supermaps and consider the set of supermaps having the same value on equivalence classes.

## II. PRELIMINARIES

Let $\mathcal{A}$ be a finite-dimensional $C^*$-algebra. Then $\mathcal{A}$ is isomorphic to a direct sum of matrix algebras, that is, there are finite-dimensional Hilbert spaces $\mathcal{H}_1, \dots \mathcal{H}_n$, such that

$$\mathcal{A} \equiv \bigoplus_j B(\mathcal{H}_j).$$

Below we always assume that $\mathcal{A}$ has this form, so that $\mathcal{A}$ is a subalgebra of block-diagonal elements in the matrix algebra $B(\mathcal{H})$, with $\mathcal{H} = \oplus_j \mathcal{H}_j$. The identity in $\mathcal{A}$ will be denoted by $I_\mathcal{A}$. We fix a trace $\mathrm{Tr}_\mathcal{A}$ on $\mathcal{A}$ to be the restriction of the trace in $B(\mathcal{H})$, we omit the subscript $\mathcal{A}$ if no confusion is possible. If $\mathcal{A} = B(\mathcal{H})$ is a matrix algebra, then we write $I_\mathcal{H}$ and $\mathrm{Tr}_\mathcal{H}$ instead of $I_{B(\mathcal{H})}$ and $\mathrm{Tr}_{B(\mathcal{H})}$. We will sometimes use the notation $\mathcal{H}_A$, $\mathcal{H}_B$, etc., for the Hilbert spaces, and $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, $\mathrm{Tr}_A = \mathrm{Tr}_{\mathcal{H}_A}$, $I_A = I_{\mathcal{H}_A}$.

If $\mathcal{B}$ is another $C^*$-algebra, then $\mathrm{Tr}_{\mathcal{A}}^{\mathcal{A}\otimes\mathcal{B}}$ will denote the partial trace on the tensor product $\mathcal{A}\otimes\mathcal{B}$, $\mathrm{Tr}_{\mathcal{A}}(a\otimes b) = \mathrm{Tr}\,(a)b$. If the input space is clear, we will sometimes denote the partial trace just by $\mathrm{Tr}_{\mathcal{A}}$.

For $a \in \mathcal{A}$, we denote by $a^T$ the transpose of $a$. Note that $\mathrm{Tr}_{\mathcal{A}}^{\mathcal{A}\otimes\mathcal{B}}(x^T) = (\mathrm{Tr}_{\mathcal{A}}^{\mathcal{A}\otimes\mathcal{B}}x)^T$ for $x \in \mathcal{A}\otimes\mathcal{B}$. If $A \subset \mathcal{A}$, then $A^T = \{a^T, a \in A\}$.

We denote by $\mathcal{A}^h$ the set of all self-adjoint elements in $\mathcal{A}$, $\mathcal{A}^+$ the convex cone of positive elements in $\mathcal{A}$ and $\mathfrak{S}(\mathcal{A})$ the set of states on $\mathcal{A}$, which will be identified with the set of density operators in $\mathcal{A}$, that is, elements $\rho \in \mathcal{A}^+$ with $\mathrm{Tr}\,\rho = 1$. If $\rho \in \mathfrak{S}(\mathcal{A})$ is invertible, then we say that $\rho$ is a faithful state. The projection onto the support of $\rho$ will be denoted by $\mathrm{supp}\,(\rho)$. If $\mathcal{A} = B(\mathcal{H})$, then we denote the set of states by $\mathfrak{S}(\mathcal{H})$. Let $\tau_{\mathcal{A}}$ denote the tracial state $t_{\mathcal{A}}^{-1}I_{\mathcal{A}}$, here $t_{\mathcal{A}} = \mathrm{Tr}\,(I_{\mathcal{A}})$. Later on, we will also need the set $\mathfrak{S}_c(\mathcal{A}) = \{a \in \mathcal{A}^+, \mathrm{Tr}\,(ca) = 1\}$ for a positive invertible element $c \in \mathcal{A}$, note that $\mathfrak{S}_{I_{\mathcal{A}}}(\mathcal{A}) = \mathfrak{S}(\mathcal{A})$.

The trace defines an inner product in $\mathcal{A}$ by $\langle a, b\rangle = \mathrm{Tr}\,(a^*b)$, with this $\mathcal{A}$ becomes a Hilbert space. If $A \subset \mathcal{A}$ is any subset, then $A^\perp$ will denote the orthogonal complement of $A$. Then $A^{\perp\perp}$ =: $[A]$ is the linear subspace, spanned by $A$. The subspace spanned by a single element $a$ will be denoted by $[a]$.

Let now $L \subseteq \mathcal{A}$ be a (complex) linear subspace. We denote by $L^h$ the set of self-adjoint elements in $L$, then $L^h$ is a real vector subspace in $\mathcal{A}^h$. The subspace $L$ is self-adjoint if $a^* \in L$ whenever $a \in L$. In this case $L = L^h \oplus iL^h$. If also $I_{\mathcal{A}} \in L$, then $L$ is called an operator system.[15] If $L$ is generated by positive elements, then we say that $L$ is positively generated. If $L_1$ and $L_2$ are subspaces in $\mathcal{A}$, then $L_1 \vee L_2$ denotes the smallest subspace containing both $L_1$ and $L_2$, and $L_1 \wedge L_2 = L_1 \cap L_2$.

### A. Channels, instruments, and POVMs

Let $\mathcal{H}$, $\mathcal{K}$ be finite-dimensional Hilbert spaces. For any linear map $T : B(\mathcal{H}) \to B(\mathcal{K})$, there is an element $X_T \in B(\mathcal{K}\otimes\mathcal{H})$, given by

$$X_T := (T \otimes id_{\mathcal{H}})(\Psi_{\mathcal{H}}), \qquad \Psi_{\mathcal{H}} = \sum_{i,j} |i\rangle\langle j| \otimes |i\rangle\langle j| \tag{1}$$

for $|i\rangle$ a canonical basis in $\mathcal{H}$. Conversely, each operator $X$ in $B(\mathcal{K}\otimes\mathcal{H})$ defines a linear map $T_X : B(\mathcal{H}) \to B(\mathcal{K})$ by

$$T_X(a) = \mathrm{Tr}_{\mathcal{H}}[(I_{\mathcal{K}} \otimes a^T)X], \qquad a \in B(\mathcal{H}). \tag{2}$$

It is easy to see that $T_{X_T} = T$ and $X_{T_X} = X$, so that the two maps are each other's inverses. The matrix $X_T$ is called the Choi matrix of $T$. We have the following:

(i) $T$ is completely positive (cp) if and only if $X_T \geq 0$.[11]
(ii) $T$ is trace-preserving if and only if $\mathrm{Tr}_{\mathcal{K}} X_T = I_{\mathcal{H}}$.

Let now $\mathcal{A} = \oplus_i B(\mathcal{H}_i)$ and $\mathcal{B} = \oplus_j B(\mathcal{K}_j)$ be finite-dimensional $C^*$-algebras. For any linear map $T : \mathcal{A} \to \mathcal{B}$ there are linear maps $T_{ij} : B(\mathcal{H}_i) \to B(\mathcal{K}_j)$ such that $T(a_i) = \oplus_j T_{ij}(a_i)$, $a_i \in B(\mathcal{H}_i)$. It is clear that $T$ is a cp map if and only if all $T_{ij}$ are cp maps. Put

$$X_T := \oplus_{i,j} X_{T_{ij}} \in \mathcal{B}\otimes\mathcal{A}. \tag{3}$$

Then it is easy to see that Eq. (2) and both (i) and (ii) hold with $\mathcal{H} = \oplus_i \mathcal{H}_i$ and $\mathcal{K} = \oplus_j \mathcal{K}_j$ (hence, we may replace $\mathrm{Tr}_{\mathcal{H}}$ and $\mathrm{Tr}_{\mathcal{K}}$ by $\mathrm{Tr}_{\mathcal{A}}$ and $\mathrm{Tr}_{\mathcal{B}}$, similarly for $I_{\mathcal{H}}$ and $I_{\mathcal{K}}$). The matrix $X_T$ is again called the Choi matrix of $T$.

Next we describe instruments and POVMs as special kinds of channels. Let $\mathcal{K}_j \equiv \mathcal{K}$ for all $j = 1, \ldots, m$, so that $\mathcal{B} = \mathbb{C}^m \otimes B(\mathcal{K})$. Then a channel $T : \mathcal{A} \to \mathcal{B}$ is called an instrument $\mathcal{A} \to B(\mathcal{K})$, with values in $\{1, \ldots, m\}$.[14] Note that $T$ is a channel if and only if $T_{ij}$ are cp maps, such that for each $i$, $T_i := \sum_j T_{ij}$ is a channel $B(\mathcal{H}_i) \to B(\mathcal{K})$. The Choi matrix of an instrument has the form $X_T = \oplus_i \sum_{j=1}^m |j\rangle\langle j| \otimes X_{ij}$, with

$$\mathrm{Tr}_{\mathcal{B}} X_T = \oplus_i \sum_j \mathrm{Tr}_{\mathcal{K}} X_{ij} = \oplus_i I_{\mathcal{H}_i} = I_{\mathcal{H}} = I_{\mathcal{A}}.$$

Let us now suppose that $\mathcal{K} = \mathbb{C}$, then $\mathcal{B}$ is the commutative $C^*$-algebra $\mathcal{B} = \mathbb{C}^m$. A channel $T : \mathcal{A} \to \mathcal{B}$ maps states onto probability distributions, hence it is given by a POVM $M_1, \dots, M_m \in \mathcal{A}^+$, $\sum_k M_k = I_{\mathcal{A}}$ as

$$T(a) = (\mathrm{Tr}\, M_1 a, \dots, \mathrm{Tr}\, M_m a). \tag{4}$$

The Choi matrix is $X_T = \sum_k |k\rangle\langle k| \otimes M_k^T$, with $\mathrm{Tr}_{\mathcal{B}} X_T = \sum_j M_j^T = I_{\mathcal{A}}$.

### *1. The link product*

Let $\mathcal{H}_i$ be Hilbert spaces, for $i = 1, 2, \dots$ and let $\mathcal{M} \subset \mathbb{N}$ be a finite set of indices. We denote $\mathcal{H}_{\mathcal{M}} := \bigotimes_{i\in\mathcal{M}} \mathcal{H}_i$. Let $\mathcal{N} \subseteq \mathbb{N}$ be another finite set and let $X \in \mathcal{H}_{\mathcal{M}}$, $Y \in \mathcal{H}_{\mathcal{N}}$ be any operators. The link product of $X$ and $Y$ was defined in Ref. 8 as the operator $X * Y \in B(\mathcal{H}_{\mathcal{M}\setminus\mathcal{N}} \otimes \mathcal{H}_{\mathcal{N}\setminus\mathcal{M}})$, given by

$$X * Y = \mathrm{Tr}_{\mathcal{M}\cap\mathcal{N}}[(I_{\mathcal{M}\setminus\mathcal{N}} \otimes Y^{T_{\mathcal{M}\cap\mathcal{N}}})(X \otimes I_{\mathcal{N}\setminus\mathcal{M}})], \tag{5}$$

where $T_{\mathcal{M}\cap\mathcal{N}}$ is the partial transpose on the space $\mathcal{H}_{\mathcal{M}\cap\mathcal{N}}$. In particular, $X * Y = X \otimes Y$, if $\mathcal{M} \cap \mathcal{N} = \emptyset$ and $X * Y = \mathrm{Tr}\,(Y^T X)$, if $\mathcal{M} = \mathcal{N}$.

*Proposition 1 (Ref. 8): The link product has the following properties.*

1. *(Associativity): Let $\mathcal{M}_i$, $i = 1, 2, 3$ be sets of indices, such that $\mathcal{M}_1 \cap \mathcal{M}_2 \cap \mathcal{M}_3 = \emptyset$. Then for $X_i \in \mathcal{H}_{\mathcal{M}_i}$,*
$$(X_1 * X_2) * X_3 = X_1 * (X_2 * X_3).$$
2. *(Commutativity): Let $X \in \mathcal{H}_{\mathcal{M}}$, $Y \in \mathcal{H}_{\mathcal{N}}$, then*
$$Y * X = E(X * Y)E,$$
*where $E$ is the unitary swap on $\mathcal{H}_{\mathcal{M}\setminus\mathcal{N}} \otimes \mathcal{H}_{\mathcal{N}\setminus\mathcal{M}}$.*
3. *(Positivity): If $X$ and $Y$ are positive, then $X * Y$ is positive.*

The interpretation of the link product is the following: If $X \in B(\mathcal{H}_1 \otimes \mathcal{H}_0)$ and $Y \in B(\mathcal{H}_2 \otimes \mathcal{H}_1)$ are the Choi matrices of maps $T_X : B(\mathcal{H}_0) \to B(\mathcal{H}_1)$ and $T_Y : B(\mathcal{H}_1) \to B(\mathcal{H}_2)$, then $X * Y$ is the Choi matrix of their composition $T_Y \circ T_X$. For $X \in B(\mathcal{H}_1)$, we have

$$Y * X = T_Y(X). \tag{6}$$

Let now $X \in \mathcal{H}_{\mathcal{M}}$ be a multipartite operator and let $\mathcal{I} \cup \mathcal{O} = \mathcal{M}$ be a partition of $\mathcal{M}$, then $X$ defines a linear map $\Phi_{X;\mathcal{I},\mathcal{O}} : \mathcal{H}_{\mathcal{I}} \to \mathcal{H}_{\mathcal{O}}$, by

$$\Phi_{X;\mathcal{I},\mathcal{O}}(a_{\mathcal{I}}) = \mathrm{Tr}_{\mathcal{H}_{\mathcal{I}}}(I_{\mathcal{H}_{\mathcal{O}}} \otimes a_{\mathcal{I}}^T)X, \qquad a_{\mathcal{I}} \in \mathcal{H}_{\mathcal{I}}. \tag{7}$$

As it was emphasized in Ref. 8, $X$ is the Choi matrix of many different maps, depending on how we choose the input and output spaces $\mathcal{I}$ and $\mathcal{O}$. The flexibility of the link product is in that it accounts for these possibilities. For example, let $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_0$ and $\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2 \cup \mathcal{M}_0$ be partitions of $\mathcal{M}$ and $\mathcal{N}$. Put $\Phi_X := \Phi_{X;\mathcal{M}_1,\mathcal{M}_0\cup\mathcal{M}_2}$ and $\Phi_Y := \Phi_{Y;\mathcal{N}_1\cup\mathcal{M}_0,\mathcal{N}_2}$. Then $X * Y$ is the Choi matrix of the map $B(\mathcal{H}_{\mathcal{M}_1\cup\mathcal{N}_1}) \to B(\mathcal{H}_{\mathcal{M}_2\cup\mathcal{N}_2})$, given by

$$\Phi_{Y*X;\mathcal{M}_1\cup\mathcal{N}_1,\mathcal{M}_2\cup\mathcal{N}_2} = (\Phi_Y \otimes id_{\mathcal{M}_2}) \circ (id_{\mathcal{N}_1} \otimes \Phi_X).$$

In the case when the input and output spaces are fixed, we will often treat a cp map and its Choi matrix as one and the same object, to shorten the discussion.

## III. EXTENSIONS OF CP MAPS AND POSITIVE FUNCTIONALS

The main goal of this paper is to study cp maps and channels from a convex subset $K$ of the state space into another $C^*$-algebra. To characterize such maps, it is crucial to know whether or when these can be extended to cp maps on the whole algebra. This section contains an extension theorem for cp maps on a vector subspace. We also prove that positive affine functionals on $K$ have positive extensions if and only if $K$ is a section, that is an intersection of the state space by a vector subspace.

### A. An extension theorem for cp maps

Let $J \subseteq \mathcal{A}$ be a subspace and let $\mathcal{K}$ be a finite-dimensional Hilbert space. Let $\mathcal{B} \subseteq B(\mathcal{K})$ be a $C^*$-algebra.

A map $\Xi : J \to \mathcal{B}$ is positive, if it maps $J \cap \mathcal{A}^+$ into the positive cone $\mathcal{B}^+$ and $\Xi$ is completely positive, if the map

$$id_{\mathcal{K}_0} \otimes \Xi : B(\mathcal{K}_0) \otimes J \to B(\mathcal{K}_0) \otimes \mathcal{B}$$

is positive, for every finite-dimensional Hilbert space $\mathcal{K}_0$. If $J$ is an operator system, that is a self-adjoint subspace containing the unit, then Arveson's extension theorem[1,15] states that any completely positive map $\Xi : J \to B(\mathcal{K})$ can be extended to a cp map $\mathcal{A} \to B(\mathcal{K})$.

The following is a consequence of this theorem in finite dimensions.

**Theorem 1:** *Let $J \subseteq \mathcal{A}$ be a self-adjoint positively generated subspace. Then any cp map $J \to \mathcal{B}$ can be extended to a cp map $\mathcal{A} \to \mathcal{B}$.*

*Proof:* Let $J^+ = J \cap \mathcal{A}^+$, so that $J$ is generated by $J^+$. There is some $\rho \in J^+$ such that the support of $\rho$ contains the supports of all other elements in $J^+$. Let us denote $p := \mathrm{supp}\,(\rho)$, then $J$ is a subspace in the algebra $\mathcal{A}_p := p\mathcal{A}p$. Denote

$$\Delta : \mathcal{A}_p \to \mathcal{A}_p, \qquad \Delta(a) = \rho^{1/2} a \rho^{1/2}.$$

Then $J' := \Delta^{-1}(J)$ is an operator system in $\mathcal{A}_p$. Moreover, $\Xi : J \to \mathcal{B}$ is a cp map if and only if $\Xi' := \Xi \circ \Delta$ is a cp map $J' \to \mathcal{B} \subseteq B(\mathcal{K})$. By Arveson's extension theorem, $\Xi'$ can be extended to a cp map $\Phi' : \mathcal{A}_p \to B(\mathcal{K})$. Let $E_{\mathcal{B}} : B(\mathcal{K}) \to \mathcal{B}$ be the trace preserving conditional expectation, then $\Phi := E_{\mathcal{B}} \circ \Phi' \circ \Delta^{-1}$ is a cp map $\mathcal{A}_p \to \mathcal{B}$ extending $\Xi$. This can be obviously extended to $\mathcal{A}$ $\square$.

### B. Sections of the state space

Let $f$ be an affine function $\mathfrak{S}(\mathcal{A}) \to \mathbb{R}^+$. Then, since $\mathfrak{S}(\mathcal{A})$ generates the positive cone $\mathcal{A}^+$, $f$ can be extended to a positive linear functional on $\mathcal{A}$. Below we discuss the possibility of such extension if $f$ is defined on some convex subset $K \subset \mathfrak{S}(\mathcal{A})$. Let us first describe a special type of such subset.

Let $K \subseteq \mathfrak{S}(\mathcal{A})$ be a convex subset and let $Q$ be the convex cone generated by $K$, then $Q = \{\lambda K, \lambda \geq 0\} \subseteq \mathcal{A}^+$. The vector subspace $[K]$ generated by $K$ is self-adjoint and $[K] = Q - Q + i(Q - Q)$.

We say that $K$ is a section of $\mathfrak{S}(\mathcal{A})$, if

$$K = [K] \cap \mathfrak{S}(\mathcal{A}). \tag{8}$$

It is clear that a section of $\mathfrak{S}(\mathcal{A})$ is convex and compact. It is also clear that (8) is equivalent with

$$Q = [K] \cap \mathcal{A}^+. \tag{9}$$

Sections of the state space can be characterized as follows.

*Proposition 2: Let $K \subset \mathfrak{S}(\mathcal{A})$ be a compact convex subset and let $Q = \{\lambda K, \lambda \geq 0\}$. Then $K$ is the section of $\mathfrak{S}(\mathcal{A})$ if and only if $a, b \in Q$, and $b \leq a$ implies $a - b \in Q$.*

*Proof:* Since we always have $Q \subseteq [K] \cap \mathcal{A}^+$, it is enough consider the inclusion $[K] \cap \mathcal{A}^+ \subseteq Q$. But $[K] \cap \mathcal{A}^+ = (Q - Q) \cap \mathcal{A}^+$ and hence any element $y \in [K] \cap \mathcal{A}^+$ has the form $y = a - b$ with $a, b \in Q$, and $b \leq a$. $\square$

*Proposition 3: Let $K \subseteq \mathfrak{S}(\mathcal{A})$. Then K is a section of $\mathfrak{S}(\mathcal{A})$ if and only if there is a subspace $J \subseteq \mathcal{A}$, such that $K = J \cap \mathfrak{S}(\mathcal{A})$.*

*Proof:* If $K$ is a section of $\mathfrak{S}(\mathcal{A})$, then we can put $J = [K]$. Conversely, let $K = J \cap \mathfrak{S}(\mathcal{A})$ for some subspace $J \subseteq \mathcal{A}$. Then $Q = J \cap \mathcal{A}^+$ and if $a, b \in Q$ with $b \leq a$, then obviously $a - b \in J \cap \mathcal{A}^+ = Q$. By Proposition 2, $K$ is a section of $\mathfrak{S}(\mathcal{A})$. □

Note that if $K = J \cap \mathfrak{S}(\mathcal{A})$ for some subspace $J$, we do not necessarily have $J = [K]$, even if $J$ is self-adjoint. The next proposition clarifies this situation.

*Proposition 4: Let $J \subseteq \mathcal{A}$ be a self-adjoint subspace and let $K = J \cap \mathfrak{S}(\mathcal{A}) \neq \emptyset$. Then there is a projection $p \in \mathcal{A}$, such that $[K] = J \cap \mathcal{A}_p$. In particular, $J = [K]$ if $J$ contains a positive invertible element.*

*Proof:* Suppose first that $J$ contains a positive invertible element $\rho$ and let $K = J \cap \mathfrak{S}(\mathcal{A})$, equivalently, $Q = J \cap \mathcal{A}^+$. Since $\mathcal{A}$ is finite-dimensional, for any $a \in J^h$, there is some $M > 0$, such that that $a \leq M\rho$, and then

$$a = M\rho - (M\rho - a) \in Q - Q.$$

This implies $J^h = Q - Q$ and since $J$ is self-adjoint, $J = [K]$.

For the general case, choose some state $\rho \in K$ such that its support contains the supports of all $\sigma \in K$, so that $K \subseteq \mathcal{A}_p$, where $p := \mathrm{supp}\,(\rho)$. Then $J_p := J \cap \mathcal{A}_p$ is a subspace in $\mathcal{A}_p$, containing the positive invertible element $\rho$ and $K = J_p \cap \mathfrak{S}(\mathcal{A}_p)$. Hence, by the first part of the proof, $[K] = J_p$. □

### C. Positive affine functions on *K*

Let $A(K)$ be the vector space of real affine functions and $A(K)^+$ the convex cone of positive affine functions over $K$. In this paragraph, we study elements in $A(K)^+$ that can be extended to a positive affine functional on $\mathfrak{S}(\mathcal{A})$, hence, are given by positive elements in $\mathcal{A}$.

Any element in $A(K)$ extends to a (unique) real linear functional on $[K]^h$ and conversely, any linear functional on $[K]^h$ defines an element in $A(K)$, so that

$$A(K) \equiv ([K]^h)^* \equiv \mathcal{A}^h|_{K^\perp} := \{a + K^\perp,\ a \in \mathcal{A}^h\}.$$

In other words, any element $\phi \in A(K)$ has the form $\phi(\sigma) = \mathrm{Tr}\, a\sigma$ for some $a \in \mathcal{A}^h$ and two elements $a_1, a_2 \in \mathcal{A}^h$ define the same $\phi \in A(K)$ if and only if $a_1 = a_2 + x$ for some $x \in K^\perp$.

Let $\pi_{K^\perp} : a \mapsto a + K^\perp$ be the quotient map. Then it is clear that $\pi_{K^\perp}(\mathcal{A}^+) \subseteq A(K)^+$. We are interested in the converse. Note that if $\bar{K}$ is the closure of $K$, then $\bar{K}$ is convex and $K^\perp = \bar{K}^\perp$, $[K] = [\bar{K}]$ and $A(K) = A(\bar{K})$, $A(K)^+ = A(\bar{K})^+$.

**Theorem 2:** *Let $K \subseteq \mathfrak{S}(\mathcal{A})$ be a nonempty convex subset. Then $A(K)^+ = \pi_{K^\perp}(\mathcal{A}^+)$ if and only if $\bar{K}$ is a section of $\mathfrak{S}(\mathcal{A})$.*

*Proof:* It is clear by the remark preceding the theorem that we may suppose that $K$ is closed.

Let $K$ be a section of $\mathfrak{S}(\mathcal{A})$, then any positive affine function on $K$ extends to a positive linear functional on $[K]$. Since positive functionals are completely positive and $[K]$ is positively generated, the assertion follows by Theorem 1.

Conversely, suppose that $K$ is not a section of $\mathfrak{S}(\mathcal{A})$. Then there is some $x \in [K] \cap \mathcal{A}^+$, such that $x \notin Q$. Since $Q$ is closed and convex, by Hahn-Banach separation theorem there is a linear functional $f$ on $\mathcal{A}^h$, such that $f(x) < s < \inf\{f(a), a \in Q\}$, for some $s \in \mathbb{R}$. This implies that $s < f(0) = 0$ and, moreover, $\lambda f(\sigma) > s$ for all $\lambda \geq 0$, $\sigma \in K$, hence $f(\sigma) \geq 0$ and $f$ defines an element $\phi \in A(K)^+$. But $\phi$ has a unique extension to $[K]$, namely, $f$ and $f(x) < s < 0$, so that $\phi$ cannot be given by an element in $\mathcal{A}^+$. □

## IV. GENERALIZED CHANNELS

Let $K \subseteq \mathfrak{S}(\mathcal{A})$ be a convex set and let $\Xi : K \to \mathcal{B}^+$ be an affine map. Then $\Xi$ extends to a linear map $[K] \to \mathcal{B}$. (Note that in general, this extension does not need to be positive.) We will say that $\Xi$ is a cp map on $K$, if this extension of $\Xi$ is completely positive. If $\Xi$ also preserves trace (equivalently, $\Xi(K) \subseteq \mathfrak{S}(\mathcal{B})$), then $\Xi$ will be called a channel on $K$.

*Remark 1:* Note that by this definition, $\Xi$ is a cp map (resp. channel) on $K$ if and only if (the extension of) $\Xi$ is a cp map (resp. channel) on $\tilde{K} := [K] \cap \mathfrak{S}(\mathcal{A})$, the smallest section of $\mathfrak{S}(\mathcal{A})$ containing $K$. Therefore, without any loss of generality we may suppose that $K$ is a section of $\mathfrak{S}(\mathcal{A})$.

**Theorem 3:** *Let $K \subseteq \mathfrak{S}(\mathcal{A})$ be a convex subset. Then any cp map on K has a cp extension to $\mathcal{A}$. If $\Phi : \mathcal{A} \to \mathcal{B}$ is a cp map, then $\Phi$ defines a channel on K if and only if its Choi matrix satisfies*

$$\mathrm{Tr}_{\mathcal{B}} X_\Phi \in I_{\mathcal{A}} + (K^T)^\perp. \tag{10}$$

*Two cp maps $\Phi_1, \Phi_2 : \mathcal{A} \to \mathcal{B}$ define the same cp map on K if and only if*

$$X_{\Phi_1} - X_{\Phi_2} \in \mathcal{B} \otimes (K^T)^\perp. \tag{11}$$

*Proof:* Since $[K]$ is positively generated, the first statement follows from Theorem 1. The map $\Phi$ defines a channel on $K$ if and only if $\mathrm{Tr}\,(\Phi(a)) = 1$ for all $a \in K$, that is,

$$\mathrm{Tr}\,(a^T) = 1 = \mathrm{Tr}\,(\Phi(a)) = \mathrm{Tr}\,((I_{\mathcal{B}} \otimes a^T)X_\Phi) = \mathrm{Tr}\,(a^T \mathrm{Tr}_{\mathcal{B}} X_\Phi), \qquad a \in K,$$

equivalently, $\mathrm{Tr}_{\mathcal{B}} X_\Phi \in I_{\mathcal{A}} + (K^T)^\perp$. Furthermore, $\Phi_1$ and $\Phi_2$ have the same value on $K$ if and only if

$$\mathrm{Tr}\,(b(\Phi_1(a) - \Phi_2(a))) = \mathrm{Tr}\,(b \otimes a^T)(X_{\Phi_1} - X_{\Phi_2}) = 0, \qquad \forall a \in K,\ b \in \mathcal{B},$$

that is, $X_{\Phi_1} - X_{\Phi_2} \in (\mathcal{B} \otimes K^T)^\perp = \mathcal{B} \otimes (K^T)^\perp$. □

Any cp map $\Phi : \mathcal{A} \to \mathcal{B}$, satisfying (10) will be called a generalized channel. Two generalized channels having the same value on $K$ will be called equivalent. If we want to stress the set $K$ (or the subspace $[K]$), we will say that $\Phi$ is a generalized channel with respect to $K$ (or $[K]$).

We will next introduce an example that will be used repeatedly throughout the paper. Let $\mathcal{A}_0$ be a finite-dimensional $C^*$-algebra and let $S : \mathcal{A} \to \mathcal{A}_0$, $T : \mathcal{A}_0 \to \mathcal{A}$ be completely positive maps. Let $J_0 \subseteq \mathcal{A}_0$ be a self-adjoint vector subspace. Then $S^{-1}(J_0) = \{a \in \mathcal{A},\ S(a) \in J_0\}$, and $T(J_0)$ are self-adjoint subspaces in $\mathcal{A}$. In particular, if $J_0 = [S(\rho)]$ is the one-dimensional subspace generated by $S(\rho)$ for some $\rho \in \mathfrak{S}(\mathcal{A})$, then $S^{-1}(J_0) \cap \mathfrak{S}(\mathcal{A})$ is the equivalence class containing $\rho$ for the equivalence relation on $\mathfrak{S}(\mathcal{A})$ induced by $S$.

*Lemma 1: Let $S : \mathcal{A} \to \mathcal{A}_0$ be a cp map and let $J_0$ be a subspace in $\mathcal{A}_0$. Then $S^{-1}(J_0)^\perp = S^*(J_0^\perp)$, where $S^* : \mathcal{A}_0 \to \mathcal{A}$ is the adjoint of S with respect to $\langle a, b\rangle = Tr\,(a^*b)$.*

*Proof:* Let $a \in \mathcal{A}$, then $\mathrm{Tr}\,(a^*S^*(b)) = \mathrm{Tr}\,(S(a^*)b) = \mathrm{Tr}\,(S(a)^*b) = 0$ for all $b \in J_0^\perp$ if and only if $S(a) \in J_0$, this implies that $S^*(J_0^\perp)^\perp = S^{-1}(J_0)$, so that $S^{-1}(J_0)^\perp = S^*(J_0^\perp)$. □

We denote by $S^T$ the linear map $\mathcal{A} \to \mathcal{A}_0$, defined by $S^T(a) = [S(a^T)]^T$. Note that the Choi matrix of $S^T$ satisfies $X_{S^T} = X_S^T$, so that $S$ is a channel if and only if $S^T$ is a channel.

*Lemma 2: Let $S : \mathcal{A} \to \mathcal{A}_0$ be a channel and let $J_0 \subseteq \mathcal{A}_0$ be a subspace. Let $J = S^{-1}(J_0)$. Then*

(i) $(J^T)^\perp = (S^T)^*((J_0^T)^\perp)$,
(ii) $I_{\mathcal{A}} + (J^T)^\perp = (S^T)^*(I_{\mathcal{A}_0} + (J_0^T)^\perp)$.

*Proof:* We have

$$S^{-1}(J_0)^T = \{a,\ S(a^T) \in J_0\} = \{a,\ S^T(a) \in J_0^T\} = (S^T)^{-1}(J_0^T),$$

(i) now follows by Lemma 1 and (ii) follows from the fact that $S^T$ is a channel, so that $(S^T)^*$ is unital. □

*Example 1 (Channels on channels):* Let $\mathcal{A} = \mathcal{B}_1 \otimes \mathcal{B}_0$, $\mathcal{A}_0 = \mathcal{B}_0$ and let $S : \mathcal{B}_1 \otimes \mathcal{B}_0 \to \mathcal{B}_0$ be the partial trace $\mathrm{Tr}_{\mathcal{B}_1}$. Let $J_0 = [I_{\mathcal{B}_0}] = \mathbb{C} I_{\mathcal{B}_0}$. The set

$$\mathcal{C}(\mathcal{B}_0, \mathcal{B}_1) := \mathrm{Tr}_{\mathcal{B}_1}^{-1}([I_{\mathcal{B}_0}]) \cap t_{\mathcal{B}_0}\mathfrak{S}(\mathcal{B}_1 \otimes \mathcal{B}_0) \tag{12}$$

is the set of all Choi matrices of channels $\mathcal{B}_0 \to \mathcal{B}_1$. Denote $J := \mathrm{Tr}_{\mathcal{B}_1}^{-1}([I_{\mathcal{B}_0}])$ and $K = J \cap \mathfrak{S}(\mathcal{B}_1 \otimes \mathcal{B}_0)$, then $K$ is a section of the state space and $\mathcal{C}(\mathcal{B}_0, \mathcal{B}_1) = t_{\mathcal{B}_0} K$. It follows that $\Xi$ is a channel on $K$ if and only if $t_{\mathcal{B}_0}^{-1} \Xi$ is a channel on $\mathcal{C}(\mathcal{B}_0, \mathcal{B}_1)$. Hence, any channel $\mathcal{C}(\mathcal{B}_0, \mathcal{B}_1) \to \mathfrak{S}(\mathcal{B})$ is given by a cp map $\Phi : \mathcal{B}_1 \otimes \mathcal{B}_0 \to \mathcal{B}$, such that $\mathrm{Tr}_{\mathcal{B}} X_\Phi \in t_{\mathcal{B}_0}^{-1} I_{\mathcal{B}_1 \otimes \mathcal{B}_2} + (K^T)^\perp$.

Since $I_{\mathcal{B}_1 \otimes \mathcal{B}_0} \in J$, we have $J = [K]$ by Proposition 4, so that $(K^T)^\perp = (J^T)^\perp$. Note also that $S^T = S$ and $S^*(a) = I_{\mathcal{B}_1} \otimes a$ for $a \in \mathcal{B}_0$. By Lemma 2,

$$(K^T)^\perp = I_{\mathcal{B}_1} \otimes [I_{\mathcal{B}_0}]^\perp$$

and taking into account that $X_\Phi \geq 0$, we get

$$\mathrm{Tr}_{\mathcal{B}} X_\Phi \in [I_{\mathcal{B}_1} \otimes (\tau_{\mathcal{B}_0} + [I_{\mathcal{B}_0}]^\perp)] \cap (\mathcal{B}_1 \otimes \mathcal{B}_0)^+ = I_{\mathcal{B}_1} \otimes \mathfrak{S}(\mathcal{B}_0). \tag{13}$$

Moreover, $\Phi_1$ and $\Phi_2$ are equivalent if and only if

$$X_{\Phi_1} - X_{\Phi_2} = I_{\mathcal{B}_1} \otimes Y, \quad Y \in \mathcal{B} \otimes \mathcal{B}_0, \ \mathrm{Tr}_{\mathcal{B}_0} Y = 0.$$ □

*Example 2 (Channels on POVMs):* Put $\mathcal{B}_1 = \mathbb{C}^m$ in example 1, then $\mathcal{C}(\mathcal{B}_0, \mathbb{C}^m)$ is the set of all POVMs on $\mathcal{B}_0$, with values in $\{1, \ldots, m\}$. If $\Phi : \mathcal{B}_1 \otimes \mathcal{B}_0 \to \mathcal{B}$ is a cp map, then the Choi matrix has the form $X_\Phi = \sum_{j=1}^m |j\rangle\langle j| \otimes X_j$, $X_j \in (\mathcal{B} \otimes \mathcal{B}_0)^+$. The condition (13) becomes

$$X_\Phi = \sum_{j=1}^m |j\rangle\langle j| \otimes X_j, \quad \mathrm{Tr}_{\mathcal{B}} X_j = \omega \ \forall j, \quad \omega \in \mathfrak{S}(\mathcal{B}_0), \tag{14}$$

and $\Phi_1$ and $\Phi_2$ are equivalent if and only if $X_{\Phi_i} = \sum_j |j\rangle\langle j| \otimes X_{ij}, \ i = 1, 2$, with

$$X_{1j} - X_{2j} = Y \ \forall j, \quad Y \in \mathcal{B} \otimes \mathcal{B}_0, \quad \mathrm{Tr}_{\mathcal{B}_0} Y = 0.$$ □

*Example 3:* Let $\mathcal{A} = B(\mathcal{H})$ and let $E = (E_1, \ldots, E_k)$ be a POVM on $B(\mathcal{H})$. Then $E$ defines a channel $S_E : B(\mathcal{H}) \to \mathbb{C}^k$ by $a \mapsto (\mathrm{Tr}\,(E_1 a), \ldots, \mathrm{Tr}\,(E_k a))$. Let $\rho$ be a faithful state and let $S_E(\rho) = \lambda = (\lambda_1, \ldots, \lambda_k)$. Let $J = S_E^{-1}([\lambda])$ and let

$$K = J \cap \mathfrak{S}(\mathcal{A}) = \{\sigma \in \mathfrak{S}(\mathcal{H}),\ \mathrm{Tr}\,(\sigma E_i) = \lambda_i, i = 1, \ldots, k\}.$$

We have $S_E^T = S_{E^T}$ and $(S_E^T)^*(x) = \sum_i x_i E_i^T$ for $x \in \mathbb{C}^k$, and since $\rho \in J$ is invertible, $(K^T)^\perp = (J^T)^\perp = S_{E^T}^*([\lambda]^\perp)$, by Lemma 2. It follows that channels $K \to \mathfrak{S}(\mathcal{B})$ are given by cp maps $\Phi : B(\mathcal{H}) \to \mathcal{B}$, such that

$$\mathrm{Tr}_{\mathcal{B}} X_\Phi = \sum_i c_i E_i^T, \qquad \sum_i c_i \lambda_i = 1.$$

Note that if $E$ is a projection valued measure, then $E^T$ is a PVM as well and positivity of $X_\Phi$ implies that we must have $c_i \geq 0$ for all $i$. Moreover, $\Phi_1$ and $\Phi_2$ are equivalent if and only if

$$X_{\Phi_1} - X_{\Phi_2} = \sum_j y_j \otimes E_j^T, \quad y_j \in \mathcal{B}, \ \sum_j \lambda_j y_j = 0.$$

More generally, let $E^i = (E_1^i, \ldots, E_{k_i}^i), i = 1, \ldots, n$ be POVMs. Put $J = \cap_i J_i$, for $J_i = S_{E^i}^{-1}([\lambda^i])$, with $\lambda_j^i = \mathrm{Tr}\,(E_j^i \rho), j = 1, \ldots, k_i, i = 1, \ldots, n$, and

$$K = J \cap \mathfrak{S}(\mathcal{A}) = \{\sigma \in \mathfrak{S}(\mathcal{A}),\ \mathrm{Tr}\,(\sigma E_j^i) = \lambda_j^i,\ j = 1, \ldots, k_i,\ i = 1, \ldots, n\}.$$

Again, $\rho \in J$, so that

$$(K^T)^\perp = (J^T)^\perp = (\cap_i J_i^T)^\perp = \vee_i (J_i^T)^\perp = \vee_i S_{E_i^T}^*([\lambda^i]^\perp).$$

It follows that channels $K \to \mathfrak{S}(\mathcal{B})$ are given by cp maps $\Phi : B(\mathcal{H}) \to \mathcal{B}$, satisfying

$$\mathrm{Tr}_{\mathcal{B}} X_\Phi = \sum_{i=1}^{n} \sum_{j=1}^{k_i} d^i_j (E^i_j)^T, \quad \sum_{i,j} d^i_j \lambda^i_j = 1,$$

and $\Phi_1$, $\Phi_2$ are equivalent if and only if

$$X_{\Phi_1} - X_{\Phi_2} = \sum_{ij} y_{ij} \otimes (E^i_j)^T, \quad y_{ij} \in \mathcal{B}, \ \sum_j y_{ij}\lambda^i_j = 0, \ \forall i. \qquad \square$$

### A. Measurements and instruments on *K*

Let $\mathcal{B} = \mathbb{C}^m \otimes B(\mathcal{K}_1)$ and let $\Phi : \mathcal{A} \to \mathcal{B}$ be a generalized channel with respect to $K$. Then there are cp maps $\Phi_j : \mathcal{A} \to B(\mathcal{K}_1), j = 1, \ldots, m$, such that $\Phi(a) = \sum_j |j\rangle\langle j| \otimes \Phi_j(a)$. Since

$$1 = \mathrm{Tr}\,(\Phi(a)) = \sum_j \mathrm{Tr}\,(\Phi_j(a)), \qquad a \in K,$$

$\sum_j \Phi_j$ is a generalized channel with respect to $K$. In this case, we will say that $\Phi$ is a generalized instrument with respect to $K$.

In particular, let $\mathcal{B} = \mathbb{C}^m$, then any cp map $\Phi : \mathcal{A} \to \mathcal{B}$ has the form (4) with some positive elements $M_j \in \mathcal{A}$ and the Choi matrix is $X_\Phi = \sum_j |j\rangle\langle j| \otimes M_j^T$. Then $\Phi$ is a generalized channel with respect to $K$ if and only if

$$\sum_j M_j = \mathrm{Tr}_{\mathcal{B}} X_\Phi^T \in I_{\mathcal{A}} + K^\perp. \tag{15}$$

Any such collection of positive operators will be called a generalized POVM (with respect to $K$). If $M$ and $N$ are generalized POVMs, then they are equivalent if and only if

$$M_j - N_j \in K^\perp, \quad \forall j. \tag{16}$$

Now let $K$ be any convex subset of $\mathfrak{S}(\mathcal{A})$. A measurement on $K$ with values in a finite set $X$ is naturally defined as an affine map from $K$ to the set of probability measures on $X$. It is clear that any generalized POVM with respect to $K$ defines a measurement on $K$ by

$$p_j(a) = \mathrm{Tr}\,(M_j a), \quad j \in X, \qquad a \in K.$$

Conversely, any measurement on $K$ is given by a collection of functions $\lambda_i \in A(K)^+$, $i \in X$, such that $\sum_i \lambda_i = 1$ (here 1 is the function identically 1 on $K$). Each $\lambda_i$ is given by some element $M_i \in \mathcal{A}^h$, such that $\sum_i M_i \in I_{\mathcal{A}} + K^\perp$. By Theorem 2, all $M_i$ can be chosen positive, and hence form a generalized POVM, if and only if $\lambda$ extends to a measurement on the section $\tilde{K}$, see Remark 1. If $K$ is a section of $\mathfrak{S}(\mathcal{A})$, then measurements on $K$ are precisely the equivalence classes of generalized POVMs. If $K$ is not a section, then Theorem 2 implies that there are measurements on $K$ that cannot be obtained by a generalized POVM.

*Example 4 (PPOVMs):* Let $\mathcal{B}_0 = \mathcal{B}(\mathcal{H}_0)$, $\mathcal{B}_1 = B(\mathcal{H}_1)$, and $\mathcal{B} = \mathbb{C}^k$ in example 1. Let us denote $\mathcal{C}(\mathcal{H}_0, \mathcal{H}_1) := \mathcal{C}(\mathcal{B}_0, \mathcal{B}_1)$ in this case. Since this is (a multiple of) a section of $\mathfrak{S}(\mathcal{H}_1 \otimes \mathcal{H}_0)$, measurements on $\mathcal{C}(\mathcal{H}_0, \mathcal{H}_1)$ are given by generalized POVMs. A collection $(M_1, \ldots, M_m)$ of operators $M_i \in B(\mathcal{H}_1 \otimes \mathcal{H}_0)^+$ is a generalized POVM with respect to $\mathcal{C}(\mathcal{H}_0, \mathcal{H}_1)$ if and only if

$$\sum_j M_j = I_{\mathcal{H}_1} \otimes \omega, \qquad \omega \in \mathfrak{S}(\mathcal{H}_0).$$

Note that these are exactly the quantum 1-testers,[5] also called process POVMs, or PPOVMs, in Ref. 16. Moreover, two PPOVMs, $M$ and $N$, are equivalent if and only if

$$M_j - N_j = I_{\mathcal{H}_1} \otimes y_j, \quad \mathrm{Tr}\,(y_j) = 0, \ \forall j.$$

Similarly, if we put $\mathcal{B}_0 = B(\mathcal{H}_0)$, $\mathcal{B}_1 = \mathbb{C}^m$, and $\mathcal{B} = \mathbb{C}^k$, we get that any measurement on the set $\mathcal{C}(B(\mathcal{H}_0), \mathbb{C}^m)$ has the form $(M_1, \ldots, M_k)$, with

$$M_j = \sum_{i=1}^{m} |i\rangle\langle i| \otimes M_{ij}, \ M_{ij} \in B(\mathcal{H}_0)^+, \quad \sum_j M_{ij} = \omega \in \mathfrak{S}(\mathcal{H}_0), \ \forall i$$

and $M$ and $N$ define the same measurement if and only if

$$M_{ij} - N_{ij} = y_j, \ \forall i, \quad \mathrm{Tr}\,(y_j) = 0, \ \forall j. \qquad \square$$

### B. Decomposition of generalized channels

Let $c \in \mathcal{A}^+$. We denote $\chi_c$: $a \mapsto c^{1/2}ac^{1/2}$. Then $\chi_c$ is a completely positive map $\mathcal{A} \to \mathcal{A}$ and $\chi_c$ defines a channel on $K$ if and only if $\mathrm{Tr}\,(\chi_c(a)) = \mathrm{Tr}\,(ac) = 1$, that is, $\mathrm{Tr}\,((I_\mathcal{A} - c)a) = 0$ for all $a \in K$. This shows that $\chi_c$ is a generalized channel if and only if

$$c \in \bigcap_{\sigma \in K} \mathfrak{S}_\sigma(\mathcal{A}) = (I_\mathcal{A} + K^\perp) \cap \mathcal{A}^+.$$

Such generalized channels with respect to $K$ will be called simple.

*Proposition 5: Let $\Phi : \mathcal{A} \to \mathcal{B}$ be a generalized channel with respect to K. Then there is a pair $(\chi, \Lambda)$, with $\chi = \chi_c$ a simple generalized channel with respect to K and $\Lambda : \mathcal{A} \to \mathcal{B}$ a channel, such that*

$$\Phi = \Lambda \circ \chi.$$

*Conversely, each such pair defines a generalized channel. If in each pair $(\chi, \Lambda)$ we take the restriction $\Lambda|_{\mathcal{A}_p}$ with $p = supp\,(c)$, then the correspondence is one-to-one.*

*Proof:* Let $\Phi : \mathcal{A} \to \mathcal{B}$ be a generalized channel. Then $\mathrm{Tr}_\mathcal{B} X_\Phi \in (I_\mathcal{A} + (K^T)^\perp) \cap \mathcal{A}^+$ or equivalently,

$$\Phi^*(I_\mathcal{B}) \in (I_\mathcal{A} + K^\perp) \cap \mathcal{A}^+.$$

Put $c = \Phi^*(I_\mathcal{B})$ and let $p = \mathrm{supp}\,(c)$. Then since $b \le \|b\| I_\mathcal{B}$ for $b \in \mathcal{B}^+$, we have $\Phi^*(b) \le \|b\| c \le \|b\|\|c\|p$. This implies that $p\Phi^*(b)p = \Phi^*(b)p = \Phi^*(b)$ for all $b \in \mathcal{B}^+$, and hence for all $b \in \mathcal{B}$, so that $\Phi^*$ maps $\mathcal{B}$ into $\mathcal{A}_p$. It follows that $\chi_{c^{-1}} \circ \Phi^*$ is well defined and unital map $\mathcal{B} \to \mathcal{A}_p$. Let $\Lambda_p$ be the adjoint map, $\Lambda_p = \Phi \circ \chi_{c^{-1}}$, then $\Lambda_p$ is a channel $\mathcal{A}_p \to \mathcal{B}$ and $\Phi = \Lambda_p \circ \chi_c$.

The channel $\Lambda_p$ can be extended to a channel $\Lambda : \mathcal{A} \to \mathcal{B}$ as

$$\Lambda(a) = \Lambda_p(a) + \omega \mathrm{Tr}\, a(1 - p), \qquad a \in \mathcal{A},$$

where $\omega \in \mathcal{B}$ is any state, and $\Phi = \Lambda \circ \chi_c$. The converse is quite obvious.

Suppose now that there are $(\chi_i, \Lambda_i)$, $i = 1, 2$, such that $\Phi_1 := \Lambda_1 \circ \chi_1 = \Lambda_2 \circ \chi_2 =: \Phi_2$. Let $\chi_i = \chi_{c_i}$. Then since $\Phi_i^*(I_\mathcal{B}) = c_i$, we have $c_1 = c_2 =: c$ and $\chi_1 = \chi_2 =: \chi$. Let $p := \mathrm{supp}\, c$. But then it is clear that if $\Lambda_i$ are defined on $\mathcal{A}_p$, then we must have $\Lambda_i = \Phi \circ \chi_c^{-1}$. $\square$

We apply this result to the set of channels on $\mathcal{C}(\mathcal{H}_0, \mathcal{H}_1)$, see example 1.

**Theorem 4:** *For any channel $\Xi : \mathcal{C}(\mathcal{H}_0, \mathcal{H}_1) \to \mathfrak{S}(\mathcal{B})$, there exists an ancillary Hilbert space $\mathcal{H}_A$, a pure state $\rho \in B(\mathcal{H}_0 \otimes \mathcal{H}_A)$ and a channel $\Lambda : B(\mathcal{H}_1 \otimes \mathcal{H}_A) \to \mathcal{B}$, such that*

$$\Xi(X_\mathcal{E}) = \Lambda \circ (\mathcal{E} \otimes id_{\mathcal{H}_A})(\rho), \qquad \mathcal{E} \in \mathcal{C}(\mathcal{H}_0, \mathcal{H}_1). \tag{17}$$

*Conversely, let $\mathcal{H}_A$ be an ancillary Hilbert space and let $\rho \in B(\mathcal{H}_0 \otimes \mathcal{H}_A)$ be a state. Let $\Lambda : B(\mathcal{H}_1 \otimes \mathcal{H}_A) \to \mathcal{B}$ be a channel. Then (17) defines a channel $\mathcal{C}(\mathcal{H}_0, \mathcal{H}_1) \to \mathfrak{S}(\mathcal{B})$.*

*Proof:* By example 1 and Proposition 5,

$$\Phi = \Lambda \circ \chi_{I_{\mathcal{H}_1} \otimes \omega}$$

with $\omega \in \mathfrak{S}(\mathcal{H}_0)$ and $\Lambda : B(\mathcal{H}_1 \otimes p\mathcal{H}_0) \to B(\mathcal{K})$ a channel, $p = \operatorname{supp}\omega$. Let now $\mathcal{E} : B(\mathcal{H}_0) \to B(\mathcal{H}_1)$ be a channel. Then we have

$$\chi_{I\otimes\omega}(X_{\mathcal{E}}) = (I_{\mathcal{H}_1} \otimes \omega^{1/2})(\mathcal{E} \otimes id_{\mathcal{H}_0})(\Psi_{\mathcal{H}_0})(I_{\mathcal{H}_1} \otimes \omega^{1/2}) = (\mathcal{E} \otimes id_{p\mathcal{H}_0})(\rho),$$

where $\rho = (I_{\mathcal{H}_0} \otimes \omega^{1/2})\Psi_{\mathcal{H}_0}(I_{\mathcal{H}_0} \otimes \omega^{1/2})$ is a pure state in $B(\mathcal{H}_0 \otimes p\mathcal{H}_0)$. Then (17) holds with $\mathcal{H}_A = p\mathcal{H}_0$.

To prove the converse, let $\mathcal{R} : B(\mathcal{H}_0) \to B(\mathcal{H}_A)$ be the cp map with Choi matrix $\rho$, then $\rho = (id_{\mathcal{H}_0} \otimes \mathcal{R})(\Psi_{\mathcal{H}_0})$. We have

$$(\mathcal{E} \otimes id_{\mathcal{H}_A})(\rho) = (\mathcal{E} \otimes id_{\mathcal{H}_A})(id_{\mathcal{H}_0} \otimes \mathcal{R})(\Psi_{\mathcal{H}_0}) = (id_{\mathcal{H}_1} \otimes \mathcal{R})(\mathcal{E} \otimes id_{\mathcal{H}_0})(\Psi_{\mathcal{H}_0}).$$

Put $\Phi = \Lambda \circ (id_{\mathcal{H}_1} \otimes \mathcal{R})$, then $\Phi$ is a cp map $B(\mathcal{H}_1 \otimes \mathcal{H}_0) \to \mathcal{B}$ and

$$\Phi^*(I_{\mathcal{B}}) = (id_{\mathcal{H}_1} \otimes \mathcal{R}^*)(I_{\mathcal{H}_1\otimes\mathcal{H}_A}) = I_{\mathcal{H}_1} \otimes \omega,$$

where $\omega = \mathcal{R}^*(I_{\mathcal{H}_A}) = \operatorname{Tr}_{\mathcal{H}_A}\rho^T$ is a state in $B(\mathcal{H}_0)$. □

Note that the analog to the above theorem for PPOVMs was proved in Ref. 16.

## V. GENERALIZED SUPERMAPS

Quantum supermaps were defined in Ref. 6 as completely positive map transforming a quantum operation to another quantum operation. More generally, supermaps on supermaps, or quantum combs, were introduced in Ref. 4. In this section, we define generalized supermaps as channels on generalized channels and show the relation to quantum combs.

Let $J \subseteq \mathcal{A}$ be a self-adjoint subspace. Denote by $\tilde{J}$ the vector subspace generated by $I_{\mathcal{A}} + (J^T)^\perp$. Then it is easy to see that $\tilde{J}$ is self-adjoint and

$$\tilde{J} = [I_{\mathcal{A}}] \vee (J^T)^\perp.$$

*Lemma 3:*

(i) *If $\rho \in J$ is any state, then*

$$(I_{\mathcal{A}} + (J^T)^\perp) \cap \mathcal{A}^+ = \tilde{J} \cap \mathfrak{S}_{\rho^T}(\mathcal{A}).$$

(ii) *If $I_{\mathcal{A}} \in J$, then $\tilde{\tilde{J}} = J$.*

(iii) *If $J = S^{-1}(J_0)$ for a channel $S : \mathcal{A} \to \mathcal{A}_0$ and a self-adjoint subspace $J_0 \subseteq \mathcal{A}_0$, then $\tilde{J} = (S^T)^*(\tilde{J}_0)$.*

*Proof:* (i) An element $x \in \tilde{J}$ has the form $x = cI_{\mathcal{A}} + x_0$, where $x_0 \in (J^T)^\perp$ and $c = \operatorname{Tr}\rho^T x$ for any state $\rho \in J$, (ii) follows from the fact that if $I_{\mathcal{A}} \in J$, then

$$\tilde{\tilde{J}} = [I_{\mathcal{A}}] \vee (\tilde{J}^T)^\perp = [I_{\mathcal{A}}] \vee ([I_{\mathcal{A}}]^\perp \wedge J) = J,$$

(iii) follows from Lemma 2. □

Let $K$ be a section of $\mathfrak{S}(\mathcal{A})$ and let $J = [K]$. We denote by $\mathcal{C}_K(\mathcal{A}, \mathcal{B})$ or $\mathcal{C}_J(\mathcal{A}, \mathcal{B})$ the set of all generalized channels $\mathcal{A} \to \mathcal{B}$ with respect to $J$. In particular, if $K = \mathfrak{S}(\mathcal{A})$, we get the set of all channels $\mathcal{C}(\mathcal{A}, \mathcal{B})$. An element $\Phi \in \mathcal{C}_J(\mathcal{A}, \mathcal{B})$ will be identified with its Choi matrix $X_\Phi \in \mathcal{B} \otimes \mathcal{A}$. In the next proposition, we characterize the set $\mathcal{C}_J(\mathcal{A}, \mathcal{B})$.

*Proposition 6: Let $K$ be a section of $\mathfrak{S}(\mathcal{A})$ and let $J = [K]$. Then*

$$\mathcal{C}_J(\mathcal{A}, \mathcal{B}) = \operatorname{Tr}_{\mathcal{B}}^{-1}(\tilde{J}) \cap \mathfrak{S}_{I_{\mathcal{B}}\otimes\rho^T}(\mathcal{B} \otimes \mathcal{A}),$$

*where $\rho$ is any element in $K$. In particular, if $K$ contains the tracial state $\tau_{\mathcal{A}}$, then $\mathcal{C}_J(\mathcal{A}, \mathcal{B}) = \operatorname{Tr}_{\mathcal{B}}^{-1}(\tilde{J}) \cap t_{\mathcal{A}}\mathfrak{S}(\mathcal{B} \otimes \mathcal{A})$.*

*Proof:* An element $X \in \mathcal{B} \otimes \mathcal{A}$ is the Choi matrix of a generalized channel with respect to $J$ if and only if $X$ is positive and

$$\mathrm{Tr}_{\mathcal{B}} X \in (I_{\mathcal{A}} + (J^T)^{\perp}) \cap \mathcal{A}^+ = \tilde{J} \cap \mathfrak{S}_{\rho^T}(\mathcal{A}),$$

by Lemma 3 (i), which is equivalent with $\mathrm{Tr}_{\mathcal{B}} X \in \tilde{J}$ and $1 = \mathrm{Tr}\,\rho^T \mathrm{Tr}_{\mathcal{B}} X = \mathrm{Tr}\,(I_{\mathcal{B}} \otimes \rho^T) X$.

If $\tau_{\mathcal{A}} \in K$, then $\mathfrak{S}_{I_{\mathcal{B}} \otimes \tau_{\mathcal{A}}^T}(\mathcal{B} \otimes \mathcal{A}) = t_{\mathcal{A}} \mathfrak{S}(\mathcal{B} \otimes \mathcal{A})$. □

This implies that if $K$ contains the tracial state, then the set of generalized channels forms a constant multiple of a section of the state space $\mathfrak{S}(\mathcal{B} \otimes \mathcal{A})$. Then any cp map that maps $C_J(\mathcal{A}, \mathcal{B})$ to another state space is a constant multiple of a generalized channel. Since the set $\mathrm{Tr}_{\mathcal{B}}^{-1}(\tilde{J})$ always contains the unit, we can repeat the process infinitely. The generalized channels obtained in this way will be called generalized supermaps.

Let $\mathcal{B}_0, \mathcal{B}_1, \mathcal{B}_2, \dots$ be finite-dimensional $C$*-algebras and let $K$ be a section of the state space $\mathfrak{S}(\mathcal{B}_0)$, such that $\tau_{\mathcal{B}_0} \in K$. Let $J = [K]$. We denote by $\mathcal{C}_J(\mathcal{B}_0, \mathcal{B}_1, \dots, \mathcal{B}_n)$ the set of all cp maps that map $\mathcal{C}_J(\mathcal{B}_0, \mathcal{B}_1, \dots, \mathcal{B}_{n-1})$ into $\mathfrak{S}(\mathcal{B}_n)$. We further introduce the following notations. Let $\mathcal{A}_n := \mathcal{B}_n \otimes \mathcal{B}_{n-1} \otimes \cdots \otimes \mathcal{B}_0$, $n = 0, 2, \dots$. Let $S_n : \mathcal{A}_n \to \mathcal{A}_{n-1}$ denote the partial trace $\mathrm{Tr}_{\mathcal{B}_n}^{\mathcal{A}_n}$, $n = 1, 2, \dots$.

**Theorem 5:** *We have for n = 1, 2, …,*

$$\mathcal{C}_J(\mathcal{B}_0, \dots, \mathcal{B}_n) = J_n \cap c_n \mathfrak{S}(\mathcal{A}_n),$$

*where*

$$J_{2k-1} = J_{2k-1}(J, \mathcal{B}_1, \dots, \mathcal{B}_{2k-1}) := S_{2k-1}^{-1}(S_{2k-2}^*(S_{2k-3}^{-1}(\dots S_1^{-1}(\tilde{J}) \dots))),$$
$$J_{2k} = J_{2k}(J, \mathcal{B}_1, \dots, \mathcal{B}_{2k}) := S_{2k}^{-1}(S_{2k-1}^*(S_{2k-2}^{-1}(\dots S_1^*(J) \dots))),$$

*and* $c_n = c_n(J, \mathcal{B}_1, \dots, \mathcal{B}_{2k-1}) := \Pi_{l=0}^{\lfloor \frac{n-1}{2} \rfloor} t_{\mathcal{B}_{n-1-2l}}$.

*Proof:* We will prove the statement by induction on $n$, together with the fact that $J_n = S_n^{-1}(\tilde{J}_{n-1})$ for $n = 1, 2, \dots$, where we put $J_0 := J$.

For $n = 1$, the statement is proved in Proposition 6 and $J_1 = S_1^{-1}(\tilde{J})$ by definition. Suppose now that this holds for some $n$. Note that since $\tilde{J}_{n-1}$ contains the unit $I_{\mathcal{A}_{n-1}}$, $J_n = S_n^{-1}(\tilde{J}_{n-1})$ contains the unit as well. Then

$$\mathcal{C}_J(\mathcal{B}_0, \dots, \mathcal{B}_{n+1}) = \frac{1}{c_n} \mathcal{C}_{J_n}(\mathcal{A}_n, \mathcal{B}_{n+1}) \tag{18}$$

and by Proposition 6,

$$\mathcal{C}_{J_n}(\mathcal{A}_n, \mathcal{B}_{n+1}) = S_{n+1}^{-1}(\tilde{J}_n) \cap t_{\mathcal{A}_n} \mathfrak{S}(\mathcal{A}_{n+1}).$$

Since $S_n^T = S_n$, we have by Lemma 3 (ii) and (iii) that

$$\tilde{J}_n = S_n^*(\tilde{\tilde{J}}_{n-1}) = S_n^*(J_{n-1}), \tag{19}$$

so that $S_{n+1}^{-1}(\tilde{J}_n) = J_{n+1}$. Finally, the proof follows from

$$\frac{t_{\mathcal{A}_n}}{c_n} = \frac{\Pi_{l=0}^n t_{\mathcal{B}_l}}{\Pi_{l=0}^{\lfloor \frac{n-1}{2} \rfloor} t_{\mathcal{B}_{n-1-2l}}} = \Pi_{l=0}^{\lfloor \frac{n}{2} \rfloor} t_{\mathcal{B}_{n-2l}} = c_{n+1}.$$

□

The above theorem can be written in the following form.

**Theorem 6:** *Let* $k := \lfloor \frac{n}{2} \rfloor$. *Then* $X \in \mathcal{C}_J(\mathcal{B}_0, \dots, \mathcal{B}_n)$ *if and only if there are positive elements* $Y^{(m)} \in \mathcal{A}_{n-2m}$ *for m = 0, …, k, such that*

$$\mathrm{Tr}_{\mathcal{B}_{n-2m}} Y^{(m)} = I_{\mathcal{B}_{n-2m-1}} \otimes Y^{(m+1)}, m = 0, \dots, k-1, \tag{20}$$

*$Y^{(0)} := X$, $Y^{(k)} \in \mathcal{C}_J(\mathcal{B}_0, \mathcal{B}_1)$ if $n = 2k + 1$ and $Y^{(k)} \in K$ if $n = 2k$.*

*Example 5 (Channels on generalized POVMs):* Let $X \in \mathcal{C}_J(\mathcal{A}, \mathbb{C}^m, \mathcal{B})$, then $X$ defines a channel on the set $\mathcal{C}_J(\mathcal{A}, \mathbb{C}^m)$ of generalized POVMs. Since $X \in \mathcal{B} \otimes \mathbb{C}^m \otimes \mathcal{A}$, we must have $X = \sum_{j=1}^m |j\rangle\langle j| \otimes X_j$, $X_j \in \mathcal{B} \otimes \mathcal{A}$. By Theorem 6, $\mathrm{Tr}_{\mathcal{B}} X = I_{\mathbb{C}^m} \otimes X_0$ for some $X_0 \in K$. It follows that if $X$ is positive,

$$X \in \mathcal{C}_J(\mathcal{A}, \mathbb{C}^m, \mathcal{B}) \iff X = \sum_{j=1}^m |j\rangle\langle j| \otimes X_j, \ \mathrm{Tr}_{\mathcal{B}} X_j = X_0 \in K, \ \forall j. \tag{21}$$

□

Note that example 2 is a special case of the above example. Another special case is the following:

*Example 6 (Channels and measurements on PPOVMs):* Let $\mathcal{H}_0, \mathcal{H}_1$ be finite-dimensional Hilbert spaces. Then $\mathcal{C}(B(\mathcal{H}_0), B(\mathcal{H}_1), \mathbb{C}^m)$ is the set of all measurements on $\mathcal{C}(\mathcal{H}_0, \mathcal{H}_1)$ with values in $\{1, \dots, m\}$, that is, the set of all PPOVMs. By (18),

$$\mathcal{C}(B(\mathcal{H}_0), \mathcal{B}(\mathcal{H}_1), \mathbb{C}^m) = \frac{1}{\dim \mathcal{H}_0} \mathcal{C}_{J_1}(B(\mathcal{H}_1 \otimes \mathcal{H}_0), \mathbb{C}^m),$$

so that

$$\mathcal{C}(B(\mathcal{H}_0), B(\mathcal{H}_1), \mathbb{C}^m, \mathcal{B}) = (\dim \mathcal{H}_0) \mathcal{C}_{J_1}(B(\mathcal{H}_1 \otimes \mathcal{H}_0), \mathbb{C}^m, \mathcal{B}),$$

here $J_1 = \mathrm{Tr}_{\mathcal{H}_1}^{-1}([I_{\mathcal{H}_0}])$. By (21), $X \in \mathcal{C}(B(\mathcal{H}_0), B(\mathcal{H}_1), \mathbb{C}^m, \mathcal{B})$ if and only if

$$X = \sum_{j=1}^m |j\rangle\langle j| \otimes X_j, \quad \mathrm{Tr}_{\mathcal{B}} X_j = X_0 \in \mathcal{C}(B(\mathcal{H}_0), B(\mathcal{H}_1)), \ \forall j.$$

Note that by Theorem 7 below, this also describes all cp maps sending POVMs with values in $\{1, \dots, m\}$ to channels $B(\mathcal{H}_0) \to \mathcal{B}$.

In particular, by putting $\mathcal{B} = \mathbb{C}^k$, we get that measurements on PPOVMs are given by collections of instruments $\Lambda_j : B(\mathcal{H}_0) \to B(\mathcal{H}_1)$ with values in $\{1, \dots, k\}$, such that their components $\Lambda_{1j}, \dots, \Lambda_{kj}$ sum to the same channel, for all $j \in \{1, \dots, m\}$.

Let now $K = \mathfrak{S}(\mathcal{B}_0)$. Then $J = \mathcal{B}_0$ and $\tilde{J} = [I_{\mathcal{B}_0}]$, so that Proposition 6 gives the usual characterization of the set $\mathcal{C}(\mathcal{B}_0, \mathcal{B}_1)$ of all Choi matrices of channels $\mathcal{B}_0 \to \mathcal{B}_1$. For $n > 1$, we have the characterization in Theorem 6 with $Y^{(k)} \in \mathfrak{S}(\mathcal{B}_0)$ if $n = 2k$ and $\mathrm{Tr}_{\mathcal{B}_1} Y^{(k)} = I_{\mathcal{B}_0}$ for $n = 2k + 1$. Suppose that all $\mathcal{B}_j$, $j = 0, 1, \dots,$ are matrix algebras, $\mathcal{B}_j = B(\mathcal{H}_j)$. Then, comparing Theorem 6 with the results in Ref. 8, we see that for $n = 2k - 1$, the set $\mathcal{C}(B(\mathcal{H}_0), \dots, B(\mathcal{H}_n))$ is precisely the set of $k$ -combs on $(\mathcal{H}_0, \dots, \mathcal{H}_{2k-1})$. We give the definition below and also give an alternative proof of the characterization of quantum combs. Note that a similar characterization was obtained in Ref. 13 for Choi matrices of strategies and co-strategies of quantum games.

### A. Quantum combs

Quantum $N$-combs were defined in Ref. 8 as a tool for description of quantum networks. A quantum 1-comb on $(\mathcal{H}_0, \mathcal{H}_1)$ is the Choi matrix of a channel $B(\mathcal{H}_0) \to B(\mathcal{H}_1)$. A quantum $N$-comb on $(\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_{2N-1})$ is the Choi matrix of a cp map, transforming $(N - 1)$-combs on $(\mathcal{H}_1, \dots, \mathcal{H}_{2N-2})$ to 1-combs on $(\mathcal{H}_0, \mathcal{H}_{2N-1})$. We use the definition of $N$-combs with the matrix algebras $B(\mathcal{H}_j)$ replaced by finite-dimensional $C^*$-algebras $\mathcal{B}_j$, $j = 0, \dots, 2N - 1$. This corresponds to conditional combs introduced in Ref. 9, which describe quantum networks with classical inputs and outputs. We show below that the $N$-combs are precisely the generalized supermaps $\mathcal{C}(\mathcal{B}_0, \dots, \mathcal{B}_{2N-1})$.

Let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ be finite-dimensional $C^*$-algebras and let $K$ be a section of $\mathfrak{S}(\mathcal{A})$, let $J = [K]$. We will describe the set of all cp maps $\mathcal{A} \to \mathcal{C} \otimes \mathcal{B}$ that transform $K$ into the set of all channels $\mathcal{B} \to \mathcal{C}$, this will be denoted by $\mathrm{Comb}_J(\mathcal{A}, \mathcal{B}, \mathcal{C})$. It will be convenient to consider this set as a subset in $\mathcal{C} \otimes \mathcal{A} \otimes \mathcal{B}$.

It is quite clear that if $X \in (\mathcal{C} \otimes \mathcal{A} \otimes \mathcal{B})^+$, then $X \in \mathrm{Comb}_J(\mathcal{A}, \mathcal{B}, \mathcal{C})$ if and only if $X * \rho \in C_J(\mathcal{A}, \mathcal{C})$ for all $\rho \in \mathfrak{S}(\mathcal{B})$, this follows from (6) and from

$$(X * \rho) * a = X * (a \otimes \rho) = (X * a) * \rho$$

for all $\rho \in \mathfrak{S}(\mathcal{B})$ and $a \in K$.

*Proposition 7: Suppose $\tau_{\mathcal{A}} \in K$. Then*

$$\mathrm{Comb}_J(\mathcal{A}, \mathcal{B}, \mathcal{C}) = \mathcal{C}_{J\otimes\mathcal{B}}(\mathcal{A} \otimes \mathcal{B}, \mathcal{C}).$$

*Proof:* Let $X$ be a positive element in $\mathcal{C} \otimes \mathcal{A} \otimes \mathcal{B}$. As we already argued above, $X \in \mathrm{Comb}_J(\mathcal{A}, \mathcal{B}, \mathcal{C})$ if and only in $X * \rho \in C_J(\mathcal{A}, \mathcal{C})$ for all $\rho \in \mathfrak{S}(\mathcal{B})$, in other words,

$$\mathrm{Tr}_{\mathcal{C}}(X * \rho) = (\mathrm{Tr}_{\mathcal{C}} X) * \rho \in \tilde{J}, \qquad \rho \in \mathfrak{S}(\mathcal{B}) \tag{22}$$

and, simultaneously,

$$\mathrm{Tr}\,(X * \rho) = \mathrm{Tr}\,(\rho^T [\mathrm{Tr}_{\mathcal{C}\otimes\mathcal{A}} X]) = t_{\mathcal{A}}, \qquad \rho \in \mathfrak{S}(\mathcal{B}), \tag{23}$$

which means that $\mathrm{Tr}_{\mathcal{C}\otimes\mathcal{A}} X = t_{\mathcal{A}} I_{\mathcal{B}}$. Moreover, we can write (22) as

$$0 = \mathrm{Tr}\,[((\mathrm{Tr}_{\mathcal{C}} X) * \rho)a] = \mathrm{Tr}\,[(\mathrm{Tr}_{\mathcal{C}} X)(a \otimes \rho^T)]$$

for all $\rho \in \mathcal{B}$ and $a \in \tilde{J}^{\perp}$, which is the same as $\mathrm{Tr}_{\mathcal{C}} X \in (\tilde{J}^{\perp} \otimes \mathcal{B})^{\perp} = \tilde{J} \otimes \mathcal{B}$. Putting this together, we get $X \in \mathrm{Comb}_J(\mathcal{A}, \mathcal{B}, \mathcal{C})$ if and only if

$$\mathrm{Tr}_{\mathcal{C}} X \in [\tilde{J} \otimes \mathcal{B}] \wedge S_{\mathcal{A}}^{-1}([I_{\mathcal{B}}]), \qquad \mathrm{Tr}\, X = t_{\mathcal{A}\otimes\mathcal{B}},$$

where $S_{\mathcal{A}} := \mathrm{Tr}_{\mathcal{A}}^{\mathcal{A}\otimes\mathcal{B}}$.

Let $Y \in \tilde{J} \otimes \mathcal{B}$, then $Y = \sum_i (t_i I_{\mathcal{A}} + x_i) \otimes b_i$, with $b_i \in \mathcal{B}$ and $x_i \in (J^T)^{\perp}$. Since $\tau_{\mathcal{A}} \in K$, we have $\mathrm{Tr}_{\mathcal{A}} Y = t_{\mathcal{A}} \sum_i t_i b_i$, so that $\mathrm{Tr}_{\mathcal{A}} Y \in [I_{\mathcal{B}}]$ if and only if $Y = c I_{\mathcal{A}\otimes\mathcal{B}} + \sum_i x_i \otimes b_i$ for some $c \in \mathbb{C}$, this implies that

$$Y \in [I_{\mathcal{A}\otimes\mathcal{B}}] \vee ((J^T)^{\perp} \otimes \mathcal{B}) = (J \otimes \mathcal{B})\check{}.$$

Conversely, let $Y \in (J \otimes \mathcal{B})\check{}$ and let $\{b_k\}_k$ be a basis in $\mathcal{B}$, such that $b_1 = I_{\mathcal{B}}$. Then there are $x_k \in (J^T)^{\perp}$, such that $Y = c I_{\mathcal{A}\otimes\mathcal{B}} + \sum_k x_k \otimes b_k = \sum_k (t_k I_{\mathcal{A}} + x_k) \otimes b_k$, with $t_1 = c$ and $t_k = 0$ for $k \neq 1$. Hence $Y \in \tilde{J} \otimes \mathcal{B}$ and clearly, $\mathrm{Tr}_{\mathcal{A}} Y \in [I_{\mathcal{B}}]$. This proves that $[\tilde{J} \otimes \mathcal{B}] \wedge S_{\mathcal{A}}^{-1}([I_{\mathcal{B}}]) = (J \otimes \mathcal{B})\check{}$, so that by Proposition 6,

$$\mathrm{Comb}_J(\mathcal{A}, \mathcal{B}, \mathcal{C}) = \mathrm{Tr}_{\mathcal{C}}^{-1}((J \otimes \mathcal{B})\check{}) \cap t_{\mathcal{A}\otimes\mathcal{B}} \mathfrak{S}(\mathcal{C} \otimes \mathcal{A} \otimes \mathcal{B}) = \mathcal{C}_{J\otimes\mathcal{B}}(\mathcal{A} \otimes \mathcal{B}, \mathcal{C}).$$

□

Let us now denote by $\mathrm{Comb}(\mathcal{B}_0, \ldots, \mathcal{B}_{2N-1})$ the set of $N$-combs.

**Theorem 7:** $\mathrm{Comb}(\mathcal{B}_0, \ldots, \mathcal{B}_{2N-1}) = \mathcal{C}(\mathcal{B}_0, \ldots, \mathcal{B}_{2N-1})$.

*Proof:* For $N = 1$, the statement is trivial. Suppose that it is true for some $N$. Let $\hat{\mathcal{A}}_{2N-1} := \mathcal{B}_{2N} \otimes \cdots \otimes \mathcal{B}_1$ and let $\hat{J}_{2N-1} := J_{2N-1}(\mathcal{B}_1, \ldots, \mathcal{B}_{2N})$ and $\hat{c}_{2N-1} = c_{2N-1}(\mathcal{B}_1, \ldots, \mathcal{B}_{2N})$, with the notations from Theorem 5. Then

$$\mathrm{Comb}(\mathcal{B}_1, \ldots, \mathcal{B}_{2N}) = \mathcal{C}(\mathcal{B}_1, \ldots, \mathcal{B}_{2N}) = \hat{J}_{2N-1} \cap \hat{c}_{2N-1} \mathfrak{S}(\hat{\mathcal{A}}_{2N-1}). \tag{24}$$

Next, let $\mathcal{A}_{2N} = \hat{\mathcal{A}}_{2N-1} \otimes \mathcal{B}_0$, $J_{2N} = J_{2N}(\mathcal{B}_0, \ldots, \mathcal{B}_{2N})$, and $c_{2N} = c_{2N}(\mathcal{B}_0, \ldots, \mathcal{B}_{2N})$. Then it is not difficult to see that $J_{2N} = \hat{J}_{2N-1} \otimes \mathcal{B}_0$ and $c_{2N} = \hat{c}_{2N-1}$. By (24) and Proposition 7,

$$\begin{aligned}
\mathrm{Comb}(\mathcal{B}_0, \ldots, \mathcal{B}_{2N+1}) &= \frac{1}{\hat{c}_{2N-1}} \mathrm{Comb}_{\hat{J}_{2N-1}}(\hat{\mathcal{A}}_{2N-1}, \mathcal{B}_0, \mathcal{B}_{2N+1}), \\
&= \frac{1}{c_{2N}} \mathcal{C}_{J_{2N}}(\mathcal{A}_{2N}, \mathcal{B}_{2N+1}), \\
&= \mathcal{C}(\mathcal{B}_0, \ldots, \mathcal{B}_{2N+1}),
\end{aligned}$$

the last equality follows from (18). □

In accordance with this result, the elements in $\mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_{2N-1})$ will be called generalized $N$-combs.

Note that an element $X \in \mathcal{C}(\mathcal{B}_0, \ldots, \mathcal{B}_{2N-1})$ is the Choi matrix of a generalized supermap $\mathcal{B}_{2N-2} \otimes \cdots \otimes \mathcal{B}_0 \to \mathcal{B}_{2N-1}$, whereas the same operator as an element in $\mathrm{Comb}(\mathcal{B}_0, \ldots, \mathcal{B}_{2N-1})$ is viewed as the Choi matrix of a cp map $\mathcal{B}_{2N-2} \otimes \cdots \otimes \mathcal{B}_1 \to \mathcal{B}_{2N-1} \otimes \mathcal{B}_0$. Note also that the set $\mathcal{C}(\mathcal{B}_0, \ldots, \mathcal{B}_{2N-1}, \mathbb{C}^k)$ is precisely the set of $N$-testers with $k$ values,[5] so that quantum testers are a special class of generalized POVMs.

### B. Decomposition of generalized supermaps

Let $k = \lfloor \frac{n}{2} \rfloor$. Let us write the algebra $\mathcal{A}_n$ as

$$\mathcal{A}_n = \mathcal{B}'_{2k} \otimes \mathcal{B}'_{2k-1} \otimes \cdots \otimes \mathcal{B}'_0, \tag{25}$$

where $\mathcal{B}'_j = \mathcal{B}_j$ for $j = 0, \ldots, n$ if $n = 2k$, and $\mathcal{B}'_j = \mathcal{B}_{j+1}$ for $j = 1, \ldots, 2k$ and $\mathcal{B}'_0 = \mathcal{B}_1 \otimes \mathcal{B}_0$ if $n = 2k + 1$. Further, let us suppose that $\mathcal{B}'_j = \oplus_{l=1}^{n_j} B(\mathcal{H}_{B^j_l})$, with minimal central projections $\{q^j_{k_j}\}$, $j = 0, 1, \ldots, 2k$. Let us denote

$$\mathcal{I}_k := \{I = (I_{2k}, \ldots, I_0) \in \mathbb{N}^{2k+1},\ \ I_j \in \{1, \ldots, n_j\},\ \ j = 0, \ldots, 2k\}$$

be the set of multi-indices. For $I \in \mathcal{I}_k$ and $l \leq k$, we denote $I^l = (I_{2l}, \ldots, I_0) \in \mathcal{I}_l$. Let $q(I) := \otimes_{l=0}^{2k} q^{2k-l}_{I_{2k-l}}$ and $\mathcal{H}_{B(I)} := \mathcal{H}_{B^{2k}_{I_{2k}} \ldots B^0_{I_0}}$, then

$$\mathcal{A}_n = \bigoplus_{I \in \mathcal{I}_k} \mathcal{H}_{B(I)}$$

and $q(I)$ are the minimal central projections in $\mathcal{A}_n$.

**Theorem 8:** *Let $X \in \mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_n)$. Let $k = \lfloor \frac{n}{2} \rfloor$. Then there are the following:*

1. *an ancillary Hilbert space $\mathcal{H}_D = \mathcal{H}_{D_0} = \mathcal{H}_{D_1} = \cdots = \mathcal{H}_{D_k}$,*
2. *elements $X_m(I^{m-1}) \in \mathcal{C}(\mathcal{B}'_{2m-1} \otimes B(\mathcal{H}_{D_{m-1}}), B(\mathcal{H}_{D_m}) \otimes \mathcal{B}'_{2m})$ for $m = 1, \ldots, k$ and for every multi-index $I \in \mathcal{I}_k$,*
3. *a state $X_0 \in B(\mathcal{H}_{D_0}) \otimes J$, if $n = 2k$ or a generalized channel $X_0 \in \mathcal{C}_J(\mathcal{B}_0, B(\mathcal{H}_{D_0}) \otimes \mathcal{B}_1)$, if $n = 2k + 1$*

*such that, for all $I \in \mathcal{I}_k$,*

$$q(I)X = I_{D_k} * X_k(I^k) * \cdots * X_1(I^1) * X_0(I_0), \tag{26}$$

*where*

$$X_m(I^m) := (I_{D_m} \otimes q^{2m}_{I_{2m}} \otimes q^{2m-1}_{I_{2m-1}} \otimes I_{D_{m-1}}) X_m(I^{m-1}),\ m = 1, \ldots, k \tag{27}$$

*and $X_0(I_0) = (I_{\mathcal{H}_{D_0}} \otimes q^0_{I_0}) X_0$.*

*Proof:* We proceed by induction on $k$. If $k = 0$, then we must have $n = 1$ and the statement is trivial. Suppose now that the theorem holds for some $k$.

Let $n$ be such that $\lfloor \frac{n}{2} \rfloor = k + 1$. Then $\mathcal{A}_n = \mathcal{B}'_{2k+2} \otimes \mathcal{B}'_{2k+1} \otimes \mathcal{A}_{n-2}$ and by Theorem 6, $X \in \mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_n)$ if and only if $X$ is positive and there is some $Y^{(1)} \in \mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_{n-2})$ such that

$$\mathrm{Tr}_{\mathcal{B}'_{2k+2}} X = I_{\mathcal{B}'_{2k+1}} \otimes Y^{(1)}.$$

Now by Theorem 11 from the Appendix, the last equation holds if and only if there is an ancillary Hilbert space $\mathcal{H}_D = \mathcal{H}_{D_k} = \mathcal{H}_{D_{k+1}}$ and

$$X_1(I_{2k+1}, \Pi_j I^k_j) \in \mathcal{C}(B(\mathcal{H}_{B^{2k+1}_{I_{2k+1}} D_k}), \mathcal{B}'_{2k+2}), \quad X_0(\Pi_j I^k_j) \in B(\mathcal{H}_{D_k B(I^k)})$$

with

$$\mathrm{Tr}_{D_k} X_0(\Pi_j I_j^k) = q(I^k) Y^{(1)} \tag{28}$$

such that

$$(I_{\mathcal{B}'_{2k+2}} \otimes q^{2k+1}_{I_{2k+1}} \otimes q(I^k))X = X_1(I_{2k+1}, \Pi_j I_j^k) * X_0(\Pi_j I_j^k)$$

for any multi-index $I \in \mathcal{I}_{k+1}$. Put

$$X_{k+1}(I^k) := \omega_{D_{k+1}} \otimes \left( \bigoplus_{i=1}^{n_{2k+1}} X_1(i, \Pi_j I_j^k) \right)$$

with an arbitrary state $\omega_{D_{k+1}} \in B(\mathcal{H}_{D_{k+1}})$. Then $X_{k+1}(I^k) \in \mathcal{C}(\mathcal{B}'_{2k+1} \otimes B(\mathcal{H}_{D_k}), B(\mathcal{H}_{D_{k+1}}) \otimes \mathcal{B}'_{2k+2})$, and

$$q(I)X = I_{D_{k+1}} * X_{k+1}(I^{k+1}) * X_0(\Pi_j I_j^k),$$

where $X_{k+1}(I^{k+1})$ is given by (27). Let now $X'_k := \bigoplus_{J \in \mathcal{I}_k} X_0(\Pi_j J_j) \in B(\mathcal{H}_{D_k}) \otimes \mathcal{A}_{n-2}$. Then by (28) and $Y^{(1)} \in \mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_{n-2})$, we get

$$\mathrm{Tr}_{B(\mathcal{H}_{D_k}) \otimes \mathcal{B}_{n-2}} X'_k = \mathrm{Tr}_{\mathcal{B}_{n-2}} Y^{(1)} = I_{\mathcal{B}_{n-3}} \otimes Y^{(2)}, \quad Y^{(2)} \in \mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_{n-4}),$$

which is equivalent with $X'_k \in \mathcal{C}_J(\mathcal{B}_0, \ldots, B(\mathcal{H}_{D_k}) \otimes \mathcal{B}_{n-2})$. Since $\lfloor \frac{n-2}{2} \rfloor = k$, we may apply the induction hypothesis to $X'_k$. Hence there is some ancilla $\mathcal{H}_E = \mathcal{H}_{E_0} = \cdots = \mathcal{H}_{E_k}$, elements $X_m(J^{m-1}) \in \mathcal{C}(\mathcal{B}'_{2m-1} \otimes B(\mathcal{H}_{E_{m-1}}), B(\mathcal{H}_{E_m}) \otimes \mathcal{B}'_{2m})$ for $m = 1, \ldots, k-1$, an element $X''_k(I^{k-1}) \in \mathcal{C}(\mathcal{B}'_{2k-1} \otimes B(\mathcal{H}_{E_{k-1}}), B(\mathcal{H}_{E_k D_k}) \otimes \mathcal{B}'_{2k})$, and $X_0 \in \mathcal{B}_0$ satisfying 3, such that for every $J \in \mathcal{I}_k$,

$$X_0(\Pi_j J_j) = q(J)X'_k = I_{E_k} * X''_k(J) * \cdots * X_0(I_0).$$

Note also that we may suppose $\mathcal{H}_E = \mathcal{H}_D$, exactly as in the proof of Theorem 11. By putting $X_k(J) = I_{E_k} * X''_k(J)$, we obtain the result. □

Theorem 8, together with Proposition 5, gives the following Corollary.

*Corollary 1: For $k \geq 1$ and for any generalized k-comb $X \in \mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_{2k-1})$, there exists a pair $(\chi, \Lambda)$, where $\chi : \mathcal{B}_0 \to \mathcal{B}_0$ is a simple generalized channel with respect to J and $X_\Lambda \in$* Comb$(\mathcal{B}_0, \ldots, \mathcal{B}_{2k-1})$*, such that*

$$\Phi_X = \Lambda \circ (id_{\mathcal{B}_{2k-1} \otimes \cdots \otimes \mathcal{B}_1} \otimes \chi).$$

*Conversely, each such pair defines an element in $\mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_{2k+1})$. In particular, $\mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_{2k+1})$ is the set of cp maps sending $\mathcal{C}(\mathcal{B}_1, \ldots, \mathcal{B}_{2k})$ to the set of generalized channels $\mathcal{C}_J(\mathcal{B}_0, \mathcal{B}_{2k+1})$.*

We will now describe how an element $Y \in \mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_{n+1})$ acts on $X \in \mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_n)$. Let $\Phi_Y : \mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_n) \to \mathcal{B}_{n+1}$ be the cp map with Choi matrix $Y$. By (6),

$$\begin{aligned}
\Phi_Y(X) = Y * X &= \mathrm{Tr}_{\mathcal{A}_n}[(I_{\mathcal{B}_{n+1}} \otimes X^T)Y], \\
&= \mathrm{Tr}_{\mathcal{A}_n}[(I_{\mathcal{B}_{n+1}} \otimes \bigoplus_I q(I)X^T) \bigoplus_{i,J} (q_i^{n+1} \otimes q(J))Y], \\
&= \mathrm{Tr}_{\mathcal{A}_n}[\bigoplus_{i,I}[(I_{\mathcal{B}_{n+1}} \otimes q(I)X^T)(q_i^{n+1} \otimes q(I))Y], \\
&= \bigoplus_i \sum_I \mathrm{Tr}_{B(I)}[(I_{\mathcal{B}_{n+1}} \otimes (q(I)X)^T)(q_i^{n+1} \otimes q(I))Y], \\
&= \bigoplus_i \sum_I ((q_i^{n+1} \otimes q(I))Y) * (q(I)X).
\end{aligned}$$

Let now $n = 2k$, so that $\lfloor\frac{n}{2}\rfloor = \lfloor\frac{n+1}{2}\rfloor = k$. Then

$$q(I)X = I_{D_k} * X_k(I^k) * \cdots * X_1(I^1) * X_0(I_0),$$
$$(q_i^{n+1} \otimes q(I))Y = I_{E_k} * Y_k(\bar{I}^k) * \cdots * Y_1(\bar{I}^1) * Y_0(\bar{I}_0).$$

Here, $\bar{I}$ is he multi-index in $\mathcal{I}_k$, such that $\bar{I}_{2k} = i$, $\bar{I}_j = I_{j+1}, j = 1, \ldots, 2k - 1$, and $\bar{I}_0 = I_0 I_1$. Then

$$((q_i^{n+1} \otimes q(I))Y) * (q(I)X) = I_{D_k E_k} * Y_k(\bar{I}^k) * X_k(I^k) * \cdots * Y_0(\bar{I}_0) * X_0(I_0),$$

this follows from Proposition 1 and 2. More explicitly, we first apply the components of the channel $Y_0(\bar{I}_0)$ to the part of $X_0(I_0)$ in $\mathcal{B}_0$, then on the part of the result in $\mathcal{B}_1$, we apply the components of the channel $X_1(I_1)$, etc., both ancillas are traced out at the end.

Similarly, if $n = 2k + 1$, then $\lfloor\frac{n+1}{2}\rfloor = k + 1$ and

$$(q_i^{n+1} \otimes q(I))Y = I_{E_{k+1}} * Y_{k+1}(\hat{I}^{k+1}) * \cdots * Y_1(\hat{I}^1) * Y_0(\hat{I}_0),$$

where $\hat{I} \in \mathcal{I}_{k+1}$ is such that $\hat{I}_{2k+2} = i$, $\hat{I}_j = I_{j-1}$ for $j = 2, \ldots 2k + 1$ and $I_0 = \hat{I}_1 \hat{I}_0$. Then

$$((q_i^{n+1} \otimes q(I))Y) * (q(I)X) = I_{D_k E_{k+1}} * Y_{k+1}(\hat{I}^{k+1}) * X_k(I^k) * \cdots * Y_1(\hat{I}_1) * X_0(I_0) * Y_0(\hat{I}_0).$$

Note that here $X_0(I_0)$ is a channel, which we apply to $Y_0(\hat{I}_0)$, etc.

*Example 7 (PPOVMs):* Let $Y \in \mathcal{C}(B(\mathcal{H}_0), B(\mathcal{H}_1), \mathbb{C}^m)$. By Theorem 8, there is some ancilla $\mathcal{H}_D$, a POVM $M(= I_{D_1} * Y_1) \in \mathcal{C}(B(\mathcal{H}_1 \otimes \mathcal{H}_D), \mathbb{C}^m)$ and a state $\rho(= Y_0) \in B(\mathcal{H}_D \otimes \mathcal{H}_0)$, such that $Y = M * \rho$. For any $X \in \mathcal{C}(\mathcal{H}_0, \mathcal{H}_1)$, we have

$$Y * X = M * X * \rho = \bigoplus_{i=1}^{m} \mathrm{Tr}\, M_i (id_D \otimes \Phi_X)(\rho),$$

where $M = (M_1, \ldots, M_m)$, compare this to Theorem 4. We will write such decomposition as $Y = (\mathcal{H}_D, (M_1, \ldots, M_m), \rho)$.

Next, let $Z \in \mathcal{C}(B(\mathcal{H}_0), B(\mathcal{H}_1), \mathbb{C}^m, B(\mathcal{H}_3) \otimes \mathbb{C}^l)$, which is the set of all instruments from PPOVMs to $B(\mathcal{H}_3)$, with values in $\{1, \ldots, l\}$. Then there is an ancilla $\mathcal{H}_E$ a channel $\xi \in \mathcal{C}(B(\mathcal{H}_0), B(\mathcal{H}_E \otimes \mathcal{H}_1))$ and an instrument $\Lambda \in \mathcal{C}(\mathbb{C}^m \otimes B(\mathcal{H}_E), B(\mathcal{H}_3) \otimes \mathbb{C}^l)$, such that

$$Z = \Lambda * \xi.$$

Here $\Lambda = \oplus_{j=1}^m \Lambda_j$, where each $\Lambda_j : B(\mathcal{H}_E) \to B(\mathcal{H}_3) \otimes \mathbb{C}^l$ is an instrument, with components $(\Lambda_{1j}, \ldots, \Lambda_{lj})$. We write $Z = (\mathcal{H}_E, (\Lambda_1, \ldots, \Lambda_m), \xi)$. Let now $Y = (\mathcal{H}_D, (M_1, \ldots, M_m), \rho)$ be a PPOVM. We have

$$\begin{aligned} Z * Y &= \bigoplus_i \sum_j \Lambda_{ij}(\mathrm{Tr}_{\mathcal{H}_D \otimes \mathcal{H}_1}(I_E \otimes M_j)(id_D \otimes \xi)(\rho)), \\ &= \bigoplus_i \sum_j \mathrm{Tr}_{\mathcal{H}_D \otimes \mathcal{H}_1}(M_j \otimes I_{\mathcal{H}_3})(id_D \otimes [(\Lambda_{ij} \otimes id_{\mathcal{H}_1}) \circ \xi])(\rho), \\ &= \bigoplus_i \sum_j \mathrm{Tr}_{\mathcal{H}_D \otimes \mathcal{H}_1}(M_j \otimes I_{\mathcal{H}_3})(id_D \otimes \hat{\Lambda}_{ij})(\rho), \end{aligned}$$

where $\hat{\Lambda}_j := (\Lambda_j \otimes id_{\mathcal{H}_1}) \circ \xi$ is an instrument $B(\mathcal{H}_0) \to B(\mathcal{H}_3 \otimes \mathcal{H}_1)$, with values in $\{1, \ldots, l\}$, such that $\sum_i \mathrm{Tr}_{\mathcal{H}_3} \circ \hat{\Lambda}_{ij} = \mathrm{Tr}_E \circ \xi$ for all $j$, compare this with example 6.

*Example 8 (Supermaps on instruments):* We next describe the set $\mathrm{Comb}(B(\mathcal{H}_0), \mathcal{B}(\mathcal{H}_1), \mathbb{C}^m \otimes B(\mathcal{H}_2), B(\mathcal{H}_3))$, that is, the set of cp maps from instruments $B(\mathcal{H}_1) \to B(\mathcal{H}_2)$ to channels $B(\mathcal{H}_0) \to B(\mathcal{H}_3)$. By Theorems 7 and 8, for any such map, there is an ancillary Hilbert space $\mathcal{H}_D$, channels $\xi_j : B(\mathcal{H}_D \otimes \mathcal{H}_2) \to B(\mathcal{H}_3), j = 1, \ldots, m$ and a channel $\xi : B(\mathcal{H}_0) \to B(\mathcal{H}_D \otimes \mathcal{H}_1)$ such that the map has the form

$$(\Lambda_1, \ldots, \Lambda_m) \mapsto \sum_j \xi_j \circ (id_D \otimes \Lambda_j) \circ \xi.$$

This seems to be more general than the supermaps considered in Ref. 6, more precisely, this map consists of $m$ supermaps in the sense of Ref. 6, which have the first channel equal to the same $\xi$. □

The decomposition given in this section can be understood as a physical realization of generalized supermaps in $\mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_n)$. It is not unique, indeed, for example, by Theorem 4, any state $\rho$ on $\mathcal{H}_0 \otimes \mathcal{H}_A$ and a POVM on $B(\mathcal{H}_1 \otimes \mathcal{H}_A)$ define a PPOVM, but (by the first part of this Theorem), we can always have a decomposition where the state is pure. The elements in $\mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_{n+1})$ do not distinguish between these different realizations, but only the generalized channels they define. We may go a step further and consider maps which recognize only the channels on $K$, defined by the generalized channels, that is, maps which give the same result on equivalent channels. This is the content of the next paragraph.

### C. Equivalence of generalized supermaps

By Theorems 3 and 5, two elements $X_1, X_2 \in \mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_n)$ are equivalent if and only if

$$X_1 - X_2 \in \mathcal{B}_n \otimes (J_{n-1}^T)^\perp. \tag{29}$$

Using Lemma 1, we get

$$(J_{n-1}^T)^\perp = S_{n-1}^*(S_{n-2}^{-1}(S_{n-3}^*(\ldots(L^T)^\perp\ldots))),$$

where $(L^T)^\perp = S_1^*([I_\mathcal{A}]^\perp \cap J)$ if $n$ is even and $(L^T)^\perp = S_1^{-1}((J^T)^\perp)$ if $n$ is odd. From this, we get the following proposition.

*Proposition 8: Let $k = \lfloor \frac{n}{2} \rfloor$. Two elements $X_1, X_2 \in \mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_n)$ are equivalent if and only if there are elements $W^{(m)} \in \mathcal{B}_n \otimes \mathcal{A}_{n-2m}$, $m = 1, \ldots k$, such that*

$$\begin{aligned}
X_1 - X_2 &= I_{\mathcal{B}_{n-1}} \otimes W^{(1)},\\
\mathrm{Tr}_{\mathcal{B}_{n-2m}} W^{(m)} &= I_{\mathcal{B}_{n-2m-1}} \otimes W^{(m+1)}, \ m = 1, \ldots, k-1,\\
W^{(k)} &\in \mathcal{B}_n \otimes J, \quad \mathrm{Tr}_{\mathcal{B}_0} W^{(k)} = 0 \qquad if\ n = 2k,\\
\mathrm{Tr}_{\mathcal{B}_1} W^{(k)} &\in \mathcal{B}_n \otimes (J^T)^\perp \qquad if\ n = 2k+1.
\end{aligned}$$

It is not clear in the present how to interpret this equivalence, in terms of the physical realizations of the channels. The next theorem gives a characterization of elements in $\mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_{n+1})$ which respect this equivalence.

**Theorem 9:** *The set of all elements in $\mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_{n+1})$ having the same value on each equivalence class of elements in $\mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_n)$ is*

$$J_{n+1} \cap (\mathcal{B}_{n+1} \otimes \mathcal{B}_n \otimes J_{n-1}) \cap c_{n+1}\mathfrak{S}(\mathcal{A}_{n+1}).$$

*In particular, if $K = \mathfrak{S}(\mathcal{B}_0)$, then this set has the form*

$$\mathcal{C}(\mathcal{B}_0, \ldots, \mathcal{B}_{n+1}) \cap \mathcal{C}(\mathcal{B}_0, \mathcal{B}_n, \mathcal{B}_{n+1}, \mathcal{B}_1, \ldots, \mathcal{B}_{n-1}), \text{ if } n \text{ is odd},$$

$$\mathcal{C}(\mathcal{B}_0, \ldots, \mathcal{B}_{n+1}) \cap \mathcal{C}(\mathcal{B}_n, \mathcal{B}_{n+1}, \mathcal{B}_0, \ldots, \mathcal{B}_{n-1}), \text{ if } n \text{ is even}.$$

*Proof:* Let $X \in \mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_{n+1})$, then it is clear from (29) that the corresponding map has the same value on equivalent elements if and only if it is equal to 0 on $\mathcal{B}_n \otimes (J_{n-1}^T)^\perp$. Equivalently,

$$0 = \mathrm{Tr}\left(b\mathrm{Tr}_{\mathcal{A}_n}[(I_{\mathcal{B}_{n+1}} \otimes Y^T)X]\right) = \mathrm{Tr}((b \otimes Y^T)X)$$

for all $b \in \mathcal{B}_{n+1}$ and $Y \in \mathcal{B}_n \otimes (J_{n-1}^T)^\perp$, that is, $X \in (\mathcal{B}_{n+1} \otimes \mathcal{B}_n \otimes J_{n-1}^\perp)^\perp = \mathcal{B}_{n+1} \otimes \mathcal{B}_n \otimes J_{n-1}$. Since $X \in \mathcal{C}_J(\mathcal{B}_0, \ldots, \mathcal{B}_{n+1})$, we get the result.

Suppose $K = \mathfrak{S}(\mathcal{B}_0)$ and let $k = \lfloor \frac{n+1}{2} \rfloor$. Since $X \in \mathcal{B}_{n+1} \otimes \mathcal{B}_n \otimes J_{n-1}$, there are positive elements $Z^{(m)} \in \mathcal{B}_{n+1} \otimes \mathcal{B}_n \otimes \mathcal{A}_{n-1-2m}$, such that

$$\mathrm{Tr}_{\mathcal{B}_{n-1-2m}} Z^{(m)} = I_{\mathcal{B}_{n-2-2m}} \otimes Z^{(m+1)}, \qquad m = 0, \ldots, k-2, \tag{30}$$

$$Z^{(k-1)} \in \mathcal{B}_{n+1} \otimes \mathcal{B}_n \otimes J \quad \text{if } n \text{ is odd}\,, \tag{31}$$

$$Z^{(k-1)} \in \mathcal{B}_{n+1} \otimes \mathcal{B}_n \otimes S_1^{-1}(\tilde{J}) \quad \text{if } n \text{ is even} \tag{32}$$

and $Z^{(0)} = X$. Suppose $n$ is odd, then by Theorem 6, we get

$$\mathrm{Tr}_{\mathcal{B}_{n+1}} \mathrm{Tr}_{\mathcal{B}_{n-1}} \dots \mathrm{Tr}_{\mathcal{B}_2} X = I_{\mathcal{B}_n \otimes \mathcal{B}_{n-2} \otimes \dots \otimes \mathcal{B}_1} \otimes Y^{(k)}$$

with $Y^{(k)} \in \mathfrak{S}(\mathcal{B}_0)$, and from (30), we have

$$\mathrm{Tr}_{\mathcal{B}_{n-1}} \mathrm{Tr}_{\mathcal{B}_{n-3}} \dots \mathrm{Tr}_{\mathcal{B}_2} X = I_{\mathcal{B}_{n-2} \otimes \mathcal{B}_{n-4} \otimes \dots \otimes \mathcal{B}_1} \otimes Z^{(k-1)}.$$

This implies $\mathrm{Tr}_{\mathcal{B}_{n+1}} Z^{(k-1)} = I_{\mathcal{B}_n} \otimes Y^{(k)}$. If $J = \mathcal{B}_0$, this together with (30) and (31) is equivalent with $X \in \mathcal{C}(\mathcal{B}_0, \mathcal{B}_n, \mathcal{B}_{n+1}, \mathcal{B}_1, \dots, \mathcal{B}_{n-1})$. Similarly, if $J = \mathcal{B}_0$ and $n$ is even, we have

$$\mathrm{Tr}_{\mathcal{B}_{n+1}} \mathrm{Tr}_{\mathcal{B}_{n-1}} \dots \mathrm{Tr}_{\mathcal{B}_1} X = I_{\mathcal{B}_n \otimes \mathcal{B}_{n-2} \otimes \dots \otimes \mathcal{B}_0}$$

and by (32), there is some positive element $Z^{(k)} \in \mathcal{B}_{n+1} \otimes \mathcal{B}_n$, such that

$$\mathrm{Tr}_{\mathcal{B}_1} Z^{(k-1)} = I_{\mathcal{B}_0} \otimes Z^{(k)}. \tag{33}$$

Then

$$\mathrm{Tr}_{\mathcal{B}_{n-1}} \mathrm{Tr}_{\mathcal{B}_{n-3}} \dots \mathrm{Tr}_{\mathcal{B}_1} X = I_{\mathcal{B}_{n-2} \otimes \mathcal{B}_{n-4} \otimes \dots \otimes \mathcal{B}_0} \otimes Z^{(k)},$$

so that we must have $\mathrm{Tr}_{\mathcal{B}_{n+1}} Z^{(k)} = I_{\mathcal{B}_n}$. This, together with (30) and (33), is equivalent with $X \in \mathcal{C}(\mathcal{B}_n, \mathcal{B}_{n+1}, \mathcal{B}_0, \dots, \mathcal{B}_{n-1})$. □

*Example 9 (Equivalence on PPOVMs):* Suppose that $Z$ is a generalized POVM on the set of PPOVMs, that is, $Z \in \mathcal{C}(B(\mathcal{H}_0), B(\mathcal{H}_1), \mathbb{C}^m, \mathbb{C}^k)$. Then by example 6, $Z = \sum_{i=1}^k \sum_{j=1}^m |i\rangle\langle i| \otimes |j\rangle\langle j| \otimes Z_{ij}$ and each $Z_{ij}$ is the Choi matrix of a cp map $\Lambda_{ij} : B(\mathcal{H}_0) \to B(\mathcal{H}_1)$, such that there is a channel $\xi$ with $\sum_j \Lambda_{ij} = \xi$ for all $i$. If $Z$ attains the same value on equivalent elements, then it defines a measurement on the set of equivalence classes of PPOVMs, that is, on the set of measurements on channels $B(\mathcal{H}_0) \to B(\mathcal{H}_1)$. By Theorem 9, this happens if and only if $Z$ is also in $\mathcal{C}(\mathbb{C}^m, \mathbb{C}^k, B(\mathcal{H}_0), B(\mathcal{H}_1))$. Using Theorem 6, we get that there are some numbers $\mu_{ij} \geq 0$, with $\sum_j \mu_{ij} = 1$ for all $i$, such that $\mathrm{Tr}_{\mathcal{H}_1} Z_{ij} = \mu_{ij} I_{\mathcal{H}_0}$. It follows that there are channels $\xi_{ij}$, such that $\Lambda_{ij} = \mu_{ij}\xi_{ij}$. We have proved the following.

For any measurement on measurements on $\mathcal{C}(B(\mathcal{H}_0), B(\mathcal{H}_1))$ with values in $\{1, \dots, m\}$, there are $\xi_{ij} \in \mathcal{C}(B(\mathcal{H}_0), B(\mathcal{H}_1))$ and numbers $\mu_{ij} \geq 0$, $\sum_j \mu_{ij} = 1$, satisfying $\sum_j \mu_{ij}\xi_{ij} = \xi$ for all $i$, such that, if a measurement on $\mathcal{C}(B(\mathcal{H}_0), B(\mathcal{H}_1))$ has an implementation $(\mathcal{H}_D, (M_1, \dots, M_m), \rho)$, then the corresponding probabilities are given by

$$p_i(\mathcal{H}_D, (M_1, \dots, M_m), \rho) = \sum_j \mu_{ij} \mathrm{Tr}\,(M_j(\xi_{ij} \otimes id_D)(\rho)).$$

Conversely, any such $\xi_{ij}$, $\mu_{ij}$ define a measurement on measurements on $\mathcal{C}(B(\mathcal{H}_0), B(\mathcal{H}_1))$. Note that if $(\mathcal{H}_D, M, \rho)$ and $(\mathcal{H}_E, N, \sigma)$ are implementations of PPOVMs, then these are equivalent if and only if $\mathrm{Tr}\,(M_j(\xi \otimes id_D)(\rho) = \mathrm{Tr}\,(N_j(\xi \otimes id_E)(\sigma)$ for any channel $\xi$. □

### D. Equivalence of combs

Any $N$-comb $X \in \mathrm{Comb}(\mathcal{B}_0, \dots, \mathcal{B}_{2N+1})$ is a cp map $\mathrm{Comb}(\mathcal{B}_1, \dots, \mathcal{B}_{2N}) \to \mathcal{B}_{2N+1} \otimes \mathcal{B}_0$. By (24) and Theorem 3, two $N$-combs $X_1$ and $X_2$ are equivalent if and only if

$$X_1 - X_2 \in \mathcal{B}_{2N+1} \otimes (\hat{J}_{2N-1}^T)^\perp \otimes \mathcal{B}_0,$$

where $\hat{J}_{2N-1} := J_{2N-1}(\mathcal{B}_1, \dots, \mathcal{B}_{2N})$.

*Proposition 9: Two elements* $X_1, X_2 \in \mathrm{Comb}(\mathcal{B}_0, \ldots, \mathcal{B}_{2N-1})$ *are equivalent if and only if there are elements* $V^{(m)} \in \mathcal{B}_{2N+1} \otimes \mathcal{A}_{2m-1}$, *m = 1, . . . , N, such that*

$$X_1 - X_2 = I_{\mathcal{B}_{2N}} \otimes V^{(N)},$$
$$\mathrm{Tr}_{\mathcal{B}_{2m-1}} V^{(m)} = I_{\mathcal{B}_{2m-2}} \otimes V^{(m-1)}, \quad m = 2, \ldots, N,$$
$$\mathrm{Tr}_{\mathcal{B}_1} V^{(1)} = 0.$$

The proof of the next theorem is the same as of Theorem 9.

**Theorem 10:** *The set elements in* $\mathrm{Comb}(\mathcal{B}_0, \ldots, \mathcal{B}_{2N+1})$ *having the same value on equivalent elements in* $\mathrm{Comb}(\mathcal{B}_1, \ldots, \mathcal{B}_{2N})$ *is equal to*

$$\mathrm{Comb}(\mathcal{B}_0, \ldots, \mathcal{B}_{2N+1}) \cap \mathrm{Comb}(\mathcal{B}_0, \mathcal{B}_1, \mathcal{B}_{2N}, \mathcal{B}_{2N+1}, \mathcal{B}_2, \ldots, \mathcal{B}_{2N-1}).$$

## VI. FINAL REMARKS

We have introduced the concept of a channel on a section of the state space of a finite-dimensional $C^*$-algebra. We proved that such channels are restrictions of completely positive maps, called generalized channels. If the section $K$ contains the tracial state, the Choi matrices of generalized channels with respect to $K$ form again a section of the state space of some $C^*$-algebra. This allows us to define generalized supermaps as completely positive maps sending generalized channels (or generalized supermaps) to states. The set of generalized supermaps is characterized as an intersection of the state space by a subspace. This might be useful, for example, in optimization problems with respect to supermaps.

Although the condition $\tau_{\mathcal{A}} \in K$ includes the most important examples of channels and combs, it might be interesting to consider supermaps for arbitrary generalized channels. By Proposition 6, this should be possible by extending our theory using the set $\mathfrak{S}_\rho(\mathcal{A})$ instead of $\mathfrak{S}(\mathcal{A})$, with an invertible element $\rho \in \mathcal{A}^+$. This can be done along similar lines.

Another possible extension of the theory is to look at the generalized channels sending a section $K_1$ to a given convex subset $K_2$ of the target state space. The set $\mathrm{Comb}_J(\mathcal{A}, \mathcal{B}, \mathcal{C})$ is a particular example of this, but arbitrary convex subset can be considered, using similar tools as were used in the present paper.

A natural question is an extension of these results to infinite dimension. For example, in the setting of the algebras of bounded operators $B(\mathcal{H})$ for infinite-dimensional Hilbert space $\mathcal{H}$, quantum supermaps were studied in Ref. 10. Channels and measurements on sections of the state space can be studied also in this case and similar results can be expected. But the identification of the set of channels with a section of a state space fails.

## ACKNOWLEDGMENTS

I would like to thank M. Ziman and T. Heinosaari for useful comments and discussions. I would also like to thank the anonymous referee for a number of valuable comments and suggestions, especially for Theorem 1 and its proof, which greatly improved the paper.

The work was supported by the grants VEGA 2/0032/09 and meta-QUTE ITMS 26240120022.

## APPENDIX: DECOMPOSITION OF SEMICAUSAL MAPS

Let $\mathcal{A} = \oplus_n B(\mathcal{H}_{A_n})$ be a finite-dimensional $C^*$-algebra and let $\mathcal{H}_A$, $\mathcal{H}_B$, $\mathcal{H}'_B$ be finite-dimensional Hilbert spaces. Let $T : \mathcal{A} \otimes B(\mathcal{H}_B) \to B(\mathcal{H}_{AB'})$ be a cp map. Then, we say that $T$ is semicausal if

$$T(I_{\mathcal{A}} \otimes b) = I_A \otimes S(b) \tag{A1}$$

for some cp map $S : B(\mathcal{H}_B) \to B(\mathcal{H}_{B'})$, and $T$ is semilocalizable, if

$$T = (id_A \otimes G) \circ (F \otimes id_B) \tag{A2}$$

for some unital cp map $F : \mathcal{A} \to B(\mathcal{H}_{AD})$ and a cp map $G : B(\mathcal{H}_{DB}) \to B(\mathcal{H}_{B'})$, where $\mathcal{H}_D$ is some (finite-dimensional) Hilbert space. The following statement was proved in Ref. 12, in the case that $\mathcal{A}$ is a matrix algebra. For the convenience of the reader, we give the modification of the proof in Ref. 12 for our slightly more general case.

*Lemma 4: Let $T : \mathcal{A} \otimes B(\mathcal{H}_B) \to \mathcal{H}_{AB'}$ be a cp map. Then T is semicausal if and only if T is semilocalizable.*

*Proof:* Any representation of $\mathcal{A} \otimes B(\mathcal{H}_B)$ has the form

$$\Pi(a \otimes b) = \oplus_n I_{E_n} \otimes a_n \otimes b = (\oplus_n I_{E_n} \otimes a_n) \otimes b$$

for some Hilbert spaces $\mathcal{H}_{E_n}$, where $a = \oplus_n a_n \in \mathcal{A}$ and $b \in B(\mathcal{H}_B)$. Hence by Stinespring representation, $T$ has the form

$$T(a \otimes b) = V^*((\oplus_n I_{E_n} \otimes a_n) \otimes b)V$$

for some linear map $V : \mathcal{H}_{AB'} \to \oplus_n \mathcal{H}_{E_n A_n B}$. Let now

$$S(b) = W^*(1_D \otimes b)W$$

be a minimal Stinespring representation of $S$. Then (A1) implies that

$$V^*(I_{\oplus_n \mathcal{H}_{E_n A_n}} \otimes b)V = (I_A \otimes W^*)(I_{AD} \otimes b)(W \otimes I_B).$$

Exactly as in Ref. 12, we get by minimality of the Stinespring representation that there is some isometry $U : \mathcal{H}_{AD} \to \oplus_n \mathcal{H}_{E_n A_n}$, such that

$$V = (U \otimes I_B)(I_A \otimes W).$$

Hence,

$$\Phi(a \otimes b) = (I_A \otimes W^*)(U^*(\oplus_n I_{E_n} \otimes a_n)U \otimes b)(I_A \otimes W),$$

so that

$$\Phi = (id_A \otimes G) \circ (F \otimes id_B) \tag{A3}$$

for the unital cp map $F : \mathcal{A} \to B(\mathcal{H}_{AD})$, given by $F(a) = U^*(\oplus_n I_{E_n} \otimes a_n)U$ and the cp map $G : B(\mathcal{H}_{DB}) \to B(\mathcal{H}'_B)$, defined as $G(d \otimes b) = W^*(d \otimes b)W$.

Conversely, if $T$ is of the form (A3), then it is clear that $T$ satisfies (A1), with

$$S(b) = G(1_D \otimes b). \tag{A4}$$

□

**Theorem 11:** *Let $\mathcal{A} = \oplus B(\mathcal{H}_{A_k})$, $\mathcal{B} = \oplus B(\mathcal{H}_{B_m})$, $\mathcal{C} = \oplus B(\mathcal{H}_{C_n})$ be finite-dimensional $C^*$-algebras, with minimal central projections $\{p_k\}_k$, $\{q_m\}_m$, and $\{r_n\}_n$, respectively. Let $X \in \mathcal{A} \otimes \mathcal{B} \otimes \mathcal{C}$ be positive. Then the following are equivalent:*

*(i) There is some positive element $Y \in \mathcal{C}$ such that*

$$\mathrm{Tr}_{\mathcal{A}} X = I_{\mathcal{B}} \otimes Y.$$

*(ii) There is an auxiliary Hilbert space $\mathcal{H}_D$, positive elements $X_0(n) \in B(\mathcal{H}_{DC_n})$ and $X_1(m, n) \in \mathcal{C}(B(\mathcal{H}_{B_m D}), \mathcal{A})$ such that*

$$X_{m,n} := (I_{\mathcal{A}} \otimes q_m \otimes r_n)X = X_1(m, n) * X_0(n).$$

*Moreover, we have*

$$\mathrm{Tr}_D X_0(n) = Y_n := r_n Y.$$

*Proof:* Suppose first that $\mathcal{B} = B(\mathcal{H}_B)$ and $\mathcal{C} = B(\mathcal{H}_C)$ are matrix algebras. We can always write $\mathcal{H}_C = \mathcal{H}_{C_1} \otimes \mathcal{H}_{C_2}$. Let us define the map $\Phi : B(\mathcal{H}_{BC_1}) \to \mathcal{A} \otimes B(\mathcal{H}_{C_2})$ by

$$\Phi(a) = X * a, \qquad a \in B(\mathcal{H}_{BC_1}).$$

Then $\Phi$ is a cp map and

$$\mathrm{Tr}_{\mathcal{A}}\Phi(a) = [\mathrm{Tr}_{\mathcal{A}} X] * a, \qquad a \in B(\mathcal{H}_{BC_1}),$$

so that $\mathrm{Tr}_{\mathcal{A}} X$ is the Choi matrix of $\mathrm{Tr}_{\mathcal{A}} \circ \Phi$. Similarly, if $\xi : B(\mathcal{H}_{C_1}) \to B(\mathcal{H}_{C_2})$ is the cp map with C-J matrix $Y$, then $I_{\mathcal{A}} \otimes Y$ is the C-J matrix of $\xi \circ \mathrm{Tr}_{\mathcal{A}}$. It follows that the maps $\Phi$ and $\xi$ satisfy

$$\mathrm{Tr}_{\mathcal{A}} \circ \Phi = \xi \circ \mathrm{Tr}_{\mathcal{A}}.$$

For the adjoints, this condition has he form $\Phi^*(I_{\mathcal{A}} \otimes c) = I_B \otimes \xi^*(c)$, for all $c \in B(\mathcal{H}_{C_2})$ which means that the map $\Phi^*$ is semicausal. By Lemma 4, (i) is equivalent with

$$\Phi = (F^* \otimes id_{C_2}) \circ (id_B \otimes G^*)$$

for a cp map $G^* : B(\mathcal{H}_{C_1}) \to B(\mathcal{H}_{DC_2})$ and a channel $F^* : B(\mathcal{H}_{BD}) \to \mathcal{A}$, with some Hilbert space $\mathcal{H}_D$. By putting $X_1$ and $X_0$ the Choi matrices of $F$ and $G$, respectively, we get (ii). Finally, (A4) implies $\mathrm{Tr}_D X_0 = Y$.

For the general case, note that $X_{m,n} \in \mathcal{A} \otimes B(\mathcal{H}_{B_m C_n})$ and

$$\mathrm{Tr}_{\mathcal{A}} X_{m,n} = (q_m \otimes r_n)\mathrm{Tr}_{\mathcal{A}} X,$$

so that (i) is equivalent with

$$\mathrm{Tr}_{\mathcal{A}} X_{m,n} = I_{B_m} \otimes Y_n, \qquad \forall m, n,$$

where $Y_n = r_n Y \in B(\mathcal{H}_{C_n})^+$. By the first part of the proof, we get that (i) holds if and only if

$$X_{m,n} = X_1'(m,n) * X_0'(m,n)$$

with positive elements $X_0'(m,n) \in B(\mathcal{H}_{D_{m,n}C_n})$, $X_1'(m,n) \in \mathcal{C}(B(\mathcal{H}_{B_m D_{m,n}}), \mathcal{A})$ for some ancillary Hilbert spaces $\mathcal{H}_{D_{m,n}}$, and such that $\mathrm{Tr}_{D_{m,n}} X_0'(m,n) = Y_n$. Note further that in the proof of Lemma 4, the cp map $G$ and the ancilla $\mathcal{H}_D$ are given by a minimal Stinespring representation of $S$. Hence $X_0'(m,n)$ and the ancilla are determined by $Y_n$, so that these depend only on $n$. Moreover, there are some $\mathcal{H}_{D_n'}$ and $\mathcal{H}_D$, such that $\mathcal{H}_D = \mathcal{H}_{D_n D_n'}$ for all $n$. Choose some state $\omega_n \in B(\mathcal{H}_{D_n'})$ for all $n$ and put

$$X_0(n) := \omega_n \otimes X_0'(n), \qquad X_1(m,n) := X_1'(m,n) \otimes I_{D_n'}.$$

Then $X_0(n) \in B(\mathcal{H}_{DC_n})$, $X_1(m,n) \in \mathcal{C}(B(\mathcal{H}_{B_m D}), \mathcal{A})$, and

$$X_1(m,n) * X_0(n) = X_1(m,n) * I_{D_n'} * \omega_n * X_0'(n) = X_1'(m,n) * X_0'(n) = X_{m,n}.$$

Clearly, also

$$\mathrm{Tr}_D X_0(n) = \mathrm{Tr}_{D_n} X_0'(n) = Y_n.$$

□

[1] Arveson, W. B., "Subalgebras of C*-algebras," Acta Math. **123**, 141–224 (1969).
[2] Bisio, A., Chiribella, G., D'Ariano, G. M., Facchini, S., and Perinotti, P., "Optimal quantum learning of a unitary transformation," Phys. Rev. A **81**, 032324 (2010).
[3] Bisio, A., D'Ariano, G. M., Perinotti, P., and Sedlak, M., J. Math. Phys. **52**, 082202 (2011).
[4] Chiribella, G., D'Ariano, G. M., and Perinotti, P., "Quantum circuit architecture," Phys. Rev. Lett. **101**, 060401 (2008).
[5] Chiribella, G., D'Ariano, G. M., and Perinotti, P., "Memory effects in quantum channel discrimination," Phys. Rev. Lett. **101**, 180501 (2008).
[6] Chiribella, G., D'Ariano, G. M., and Perinotti, P., "Transforming quantum operations: Quantum supermaps," Europhys. Lett. **83**, 30004 (2008).
[7] Chiribella, G., D'Ariano, G. M., and Perinotti, P., "Optimal cloning of unitary transformations," Phys. Rev. Lett. **101**, 180504 (2008).
[8] Chiribella, G., D'Ariano, G. M., and Perinotti, P., "Theoretical framework for quantum networks," Phys. Rev. A **80**, 022339 (2009).

[9] Chiribella, G., D'Ariano, G. M., Perinotti, P., Schlingemann, D. M., and Werner, R. F., e-print arXiv:0905.3801.
[10] Chiribella, G., Toigo, A., and Umanità, V., e-print arXiv:1012.3197v2.
[11] Choi, M. D., "Completely positive maps on complex matrices," Linear Algebr. Appl. **10**, 285–290 (1975).
[12] Eggeling, T., Schlingemann, D., and Werner, R. F., "Semicausal operations are semilocalizable," Europhys. Lett. **57**, 782–788 (2002).
[13] Gutoski, G. and Watrous, J., *Proceedings of the 39th ACM Symposium on Theory of Computing*, (STOC 2007) 565–574 2007.
[14] Holevo, A. S., *Statistical Structure of Quantum Theory* (Springer-Verlag, Berlin, 2001).
[15] Paulsen, V., *Completely Bounded Maps and Operator Algebras* (Cambridge University Press, Cambridge, England, 2003).
[16] Ziman, M., "Process POVM: A mathematical framework for the description of process tomography experiments," Phys. Rev. A **77**, 062112 (2008).

JOURNAL OF MATHEMATICAL PHYSICS **55**, 022201 (2014)

# Base norms and discrimination of generalized quantum channels

A. Jenčová
*Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, Bratislava, Slovakia*

(Received 17 September 2013; accepted 16 January 2014;
published online 12 February 2014)

We introduce and study norms in the space of hermitian matrices, obtained from base norms in positively generated subspaces. These norms are closely related to discrimination of so-called generalized quantum channels, including quantum states, channels, and networks. We further introduce generalized quantum decision problems and show that the maximal average payoffs of decision procedures are again given by these norms. We also study optimality of decision procedures, in particular, we obtain a necessary and sufficient condition under which an optimal 1-tester for discrimination of quantum channels exists, such that the input state is maximally entangled.
 [http://dx.doi.org/10.1063/1.4863715]

## I. INTRODUCTION AND PRELIMINARIES

It is well known that in the problem of discrimination of quantum states, the best possible distinguishability of two states $\rho_0$ and $\rho_1$ is given by the trace norm $\|\rho_0 - \rho_1\|_1$.[10,11] The set of states forms a base of the convex cone of positive operators and the restriction of the trace norm to hermitian operators is the corresponding base norm. Similarly, it was shown in Ref. 20 that more general distinguishability measures, obtained by specification of the allowed measurements, e.g., for bipartite states, are obtained from base norms associated with more general positive cones. This correspondence is related to duality of the base norm and the order unit norm, with respect to a given positive cone.

In a similar problem for quantum channels, and recently also quantum networks, the diamond norm $\|\cdot\|_\diamond$ for channels,[14] resp. the strategy $N$-norm $\|\cdot\|_{N\diamond}$[9,3] for networks is obtained. Via the Choi isomorphism, quantum networks are represented by certain positive operators on the tensor product of the input and output spaces, so-called $N$-combs,[2,4] see also Ref. 8. The set of $N$-combs is the intersection of the multipartite state space by a positively generated subspace of the real vector space of hermitian operators. Since this subspace inherits the order structure and the set of $N$-combs forms a base of its positive cone, it is natural to expect that the distinguishability norm $\|\cdot\|_{N\diamond}$ is in fact the corresponding base norm.

Motivated by this question, we study positively generated subspaces of the space of hermitian operators $B_h(\mathcal{H})$ acting on a finite dimensional Hilbert space $\mathcal{H}$. For a given base $B$ of the positive cone, we define a distinguishability measure in terms of tests that are defined as affine maps $B \to [0, 1]$ and show that this measure is given by the base norm. This, in fact, is easy to see for any finite dimensional ordered vector space. We then study a natural extension of this norm to $B_h(\mathcal{H})$ and its dual norm. An example of such a base is the set of Choi matrices of so-called generalized channels. The set of $N$-combs is a special case. For $N$-combs, the obtained norm coincides with $\|\cdot\|_{N\diamond}$ and we recover some of the results of Ref. 9 concerning the dual norm. Moreover, we find a suitable expression for this norm, closely related to the definition of $\|\cdot\|_\diamond$.

In Sec. IV, we introduce generalized quantum decision problems with respect to a base $B$. We show that the maximal average payoff (or minimal average loss) of a generalized decision procedure is again given by a base norm. We find optimality conditions for generalized decision procedures, in particular, for quantum measurements and testers. In the case of multiple hypothesis testing for states, we get the results obtained previously in Refs. 17, 7, and 12. In the case of discrimination of quantum channels, we find a necessary and sufficient condition for existence of an optimal tester such that the input state is maximally entangled.

The rest of the present section contains some basic definitions and preliminary results on discrimination of quantum devices, as well as convex cones, bases, and base norms.

### A. Discrimination of quantum states, channels, and networks

Let $\mathcal{H}$ be a finite dimensional Hilbert space and let $B(\mathcal{H})$ be the set of bounded operators on $\mathcal{H}$. We denote by $B_h(\mathcal{H})$ the set of self-adjoint operators, $B(\mathcal{H})^+$ the cone of positive operators and $\mathfrak{S}(\mathcal{H}) := \{\rho \geq 0, \operatorname{Tr}\rho = 1\}$ the set of states in $B(\mathcal{H})$. We will also use the notation $B(\mathcal{H})^{++}$ for the set of strictly positive elements in $B(\mathcal{H})$. Let $\mathcal{K}$ be another finite dimensional Hilbert space. It is well known that $B(\mathcal{K}\otimes\mathcal{H})$ corresponds to the set of all linear maps $B(\mathcal{H}) \to B(\mathcal{K})$, via the Choi representation:[1]

$$X_\Phi = (\Phi \otimes id_{\mathcal{H}})(\Psi), \qquad \Phi_X(a) = \operatorname{Tr}_{\mathcal{H}}[(I_{\mathcal{K}} \otimes a^{\mathsf{T}})X], \tag{1}$$

here $\Psi = |\psi\rangle\langle\psi|$ and $|\psi\rangle = \sum_i |i\rangle \otimes |i\rangle$ for an orthonormal basis (ONB) $\{|i\rangle, i = 1, \dots, \dim(\mathcal{H})\}$ in $\mathcal{H}$, $a^{\mathsf{T}}$ denotes transpose of $a$ with respect to this basis. In this correspondence, $B(\mathcal{K}\otimes\mathcal{H})^+$ is identified with the set of completely positive maps and $B_h(\mathcal{K}\otimes\mathcal{H})$ with hermitian maps, that is, maps satisfying $\Phi(a^*) = \Phi(a)^*$.

Consider the problem of quantum state discrimination: suppose the quantum system represented by $\mathcal{H}$ is known to be in one of two given states $\rho_0$ or $\rho_1$ and the task is to decide which of them is the true state. This is done by using a test, that is a binary positive operator valued measure (POVM). This is given by an operator $0 \leq M \leq I$, with the interpretation that $\operatorname{Tr} M\rho$ is the probability of deciding for $\rho_0$ if the true value of the state is $\rho$. Equivalently, a test can be defined as an affine map $\mathfrak{S}(\mathcal{H}) \to [0, 1]$.

Given an *a priori* probability $0 \leq \lambda \leq 1$ that the true state is $\rho_0$, we need to minimize the average probability of error over all tests, that is to find the value of

$$\Pi_\lambda(\rho_0, \rho_1) := \min_{0\leq M\leq I} \lambda \operatorname{Tr}(I - M)\rho_0 + (1-\lambda)\operatorname{Tr} M\rho_1,$$

this is the minimum Bayes error probability. Then[10,11]

$$\Pi_\lambda(\rho_0, \rho_1) = \frac{1}{2} - \frac{1}{2}\|\lambda\rho_0 - (1-\lambda)\rho_1\|_1,$$

where $\|a\|_1 := \operatorname{Tr}|a|$, $a \in B(\mathcal{H})$ is the trace norm.

Let now $\mathcal{H}$ and $\mathcal{K}$ be two finite dimensional Hilbert spaces and consider the problem of discrimination of channels. Here we have to decide between two channels $\Phi_0$ and $\Phi_1$ and this time the tests are given by binary quantum 1-testers,[3] or PPOVMs,[21] which are positive operators $T \in B(\mathcal{K}\otimes\mathcal{H})^+$, such that $T \leq I_{\mathcal{K}} \otimes \sigma$ for some $\sigma \in \mathfrak{S}(\mathcal{H})$. These correspond to triples $(\mathcal{H}_A, \rho, M)$, where $\mathcal{H}_A$ is an ancillary Hilbert space, $\rho \in \mathfrak{S}(\mathcal{H}\otimes\mathcal{H}_A)$ and $0 \leq M \leq I$, $M \in B(\mathcal{K}\otimes\mathcal{H}_A)$. The probability of choosing $\Phi_0$ if the true value is $\Phi$ for a tester $T$ is given by

$$p(T, \Phi) := \operatorname{Tr} T X_\Phi = \operatorname{Tr} M(\Phi \otimes id_A)(\rho).$$

The minimum Bayes error probability is now

$$\Pi^1_\lambda(\Phi_0, \Phi_1) := \min_T \lambda(1 - p(T, \Phi_0)) + (1-\lambda)p(T, \Phi) = \frac{1}{2} - \frac{1}{2}\|\lambda\Phi_0 - (1-\lambda)\Phi_1\|_\diamond,$$

where the diamond norm $\|\Phi\|_\diamond$ for a hermitian map $\Phi$ is defined as[14,19]

$$\|\Phi\|_\diamond = \sup_{\dim(\mathcal{L}')<\infty} \sup_{\rho\in\mathfrak{S}(\mathcal{H}\otimes\mathcal{L}')} \|\Phi \otimes id_{\mathcal{L}'}(\rho)\|_1$$
$$= \sup_{\rho\in\mathfrak{S}(\mathcal{H}\otimes\mathcal{L})} \|\Phi \otimes id_{\mathcal{L}}(\rho)\|_1, \qquad \dim(\mathcal{L}) = \dim(\mathcal{H}).$$

By duality, this norm is related to the *cb*-norm for completely bounded linear maps, see Ref. 16.

Let now $\{\mathcal{H}_0, \mathcal{H}_1, \dots, \mathcal{H}_{2N-1}\}$ be finite dimensional Hilbert spaces. Consider a sequence of channels $\Phi_i : B(\mathcal{H}_{2i-2}\otimes\mathcal{H}_A) \to B(\mathcal{H}_{2i-1}\otimes\mathcal{H}_A)$, $i = 1, \dots, N$, connected by the ancilla $\mathcal{H}_A$ as

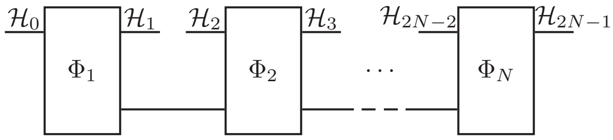


FIG. 1. A deterministic quantum $N$-comb.

indicated in Fig. 1 (the first and last ancilla are traced out). This defines a channel $\Phi : B(\mathcal{H}_0 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_{2N-2}) \to B(\mathcal{H}_1 \otimes \mathcal{H}_3 \otimes \cdots \otimes \mathcal{H}_{2N-1})$, such channels describe quantum networks. The channels $\Phi_1, \ldots, \Phi_N$ are not unique, in fact, these can always be supposed to be isometries. A (deterministic) quantum $N$-comb is defined as the Choi matrix $X_\Phi$ of such a channel, see Ref. 4 for more about quantum networks and $N$-combs. The same definition, called a (non-measuring) quantum $N$-round strategy, was also introduced in Ref. 8. A (non-measuring) quantum $N$-round co-strategy can be defined as an $(N + 1)$-strategy for the sequence of spaces $\{\mathbb{C}, \mathcal{H}_0, \ldots, \mathcal{H}_{2N-1}, \mathbb{C}\}$.

The tests for discrimination of two networks $\Phi^0$ and $\Phi^1$ are given by quantum $N$-testers, which are obtained by an $(N + 1)$-comb such that the first channel has 1-dimensional input space (hence is a state) and a (binary) POVM is applied to the ancilla,[4,3] see Figs. 2 and 3. This can be represented by a pair $(T_0, T_1)$ of positive operators, such that $T_0 + T_1$ is an $(N + 1)$-round co-strategy.[4,8,9]

The minimal Bayes error probability now has the form

$$\Pi^N_\lambda(\Phi^0, \Phi^1) = \frac{1}{2} - \frac{1}{2}\|\lambda\Phi^0 - (1-\lambda)\Phi^1\|_{N\diamond},$$

where the norm $\|\cdot\|_{N\diamond}$ was introduced in Ref. 3 as

$$\|\Phi\|_{N\diamond} = \sup_T \|(T_0 + T_1)^{1/2} X_\Phi (T_0 + T_1)^{1/2}]\|_1, \tag{2}$$

for any hermitian $\Phi$. Another expression for this norm was found in Ref. 9:

$$\|\Phi\|_{N\diamond} = \sup_T \mathrm{Tr}\, X_\Phi (T_0 - T_1). \tag{3}$$

In both cases, the supremum is taken over all $N$-testers. The dual norm was also obtained in Ref. 9 as

$$\|\Phi\|^*_{N\diamond} = \sup_S \mathrm{Tr}\, X_\Phi (S_1 - S_0),$$

where the supremum is taken over the set of pairs of positive operators such that $S_0 + S_1$ is an $N$-round strategy ($N$-comb).

### B. Convex cones, bases, and base norms

Let $\mathcal{V}$ be a finite dimensional real vector space and let $\mathcal{V}^*$ be the dual space, with duality $\langle\cdot, \cdot\rangle$. A subset $Q \subset \mathcal{V}$ is a convex cone if $\lambda q_1 + \mu q_2 \in Q$ whenever $q_1, q_2 \in Q$ and $\lambda, \mu \geq 0$. The cone is pointed if $Q \cap -Q = \{0\}$ and generating if $\mathcal{V} = Q - Q$. Closed pointed convex cones are in one-to-one correspondence with partial orders in $\mathcal{V}$, by $x \leq_Q y \Leftrightarrow y - x \in Q$.

The dual cone of $Q$ is defined as

$$Q^* = \{f \in \mathcal{V}^*, \langle f, q\rangle \geq 0, q \in Q\}.$$

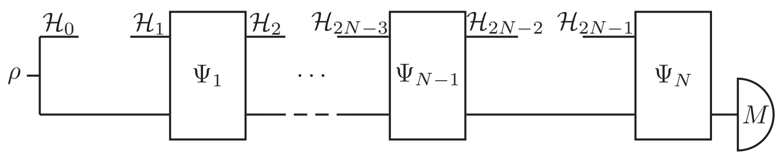


FIG. 2. A quantum $N$-tester.

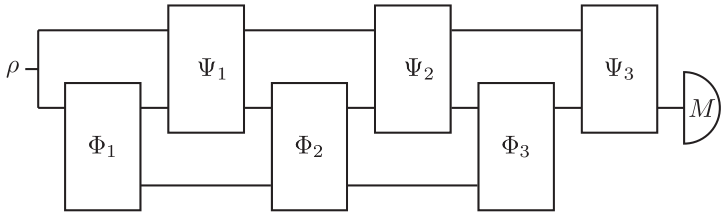


FIG. 3. A 3-tester $\Psi$ applied to a 3-comb $\Phi$.

This is a closed convex cone and $Q^{**} = Q$ if $Q$ is closed. Moreover, a closed convex cone $Q$ is pointed if and only if $Q^*$ is generating . A closed pointed generating convex cone is called a proper cone.

A base of the proper cone $Q$ is a compact convex subset $B \subset Q$, such that each nonzero element $q \in Q$ has a unique representation in the form $q = \lambda b$ with $b \in B$ and $\lambda > 0$. It is clear that any base generates the cone $Q$, in the sense that $Q = \bigcup_{\lambda \geq 0} \lambda B$. Then any element $v \in \mathcal{V}$ can be written as $v = \lambda b_1 - \mu b_2$, $\lambda, \mu \geq 0$, $b_1, b_2 \in B$.

For any base $B$, the map $Q \ni q = \lambda b \mapsto \lambda$ extends uniquely to a linear functional $e_B \in Q^*$ and we have $B = \{q \in Q, \langle e_B, q\rangle = 1\}$.

*Lemma 1. Let $f \in Q^*$. Then $f \in int(Q^*)$ if and only if*

$$B_f := \{q \in Q, \langle f, q\rangle = 1\}$$

*is a base of $Q$.*

*Proof.* It is quite clear that $B_f$ is a base of $Q$ if and only if $\langle f, q\rangle > 0$ for any nonzero $q \in Q$. By Theorem 11.6 of Ref. 18, this is equivalent with $f \in int(Q^*)$. □

Let $\leq$ denote the order in $\mathcal{V}$ given by $Q$. An element $e \in \mathcal{V}$ is an order unit in $\mathcal{V}$ if for any $v \in \mathcal{V}$, there is some $r > 0$ such that $re \geq v$. It is easy to see that $e$ is an order unit if and only if $e \in int(Q)$. Consequently,

*Corollary 1. Any base $B$ of $Q$ defines an order unit $e_B$ in $\mathcal{V}^*$ and, conversely, any order unit $e$ in $\mathcal{V}^*$ defines a base $B_e$ of $Q$. We have $e_{B_e} = e$ and $B_{e_B} = B$.*

Let $B$ be a base of $Q$. The corresponding base norm in $\mathcal{V}$ is defined by

$$\|v\|_B = \inf\{\lambda + \mu,\ v = \lambda b_1 - \mu b_2, \lambda, \mu \geq 0,\ b_1, b_2 \in B\}.$$

It is clear that $\|q\|_B = \langle e_B, q\rangle$ for all $q \in Q$. Let $\mathcal{V}_1$ be the unit ball of $\|\cdot\|_B$ in $\mathcal{V}$, then

$$\mathcal{V}_1 = \{\lambda b_1 - \mu b_2,\ b_1, b_2 \in B, \lambda, \mu \geq 0, \lambda + \mu = 1\} = co(B \cup -B),$$

where $co(A)$ denotes the convex hull of $A \subset \mathcal{V}$. Let $\|\cdot\|_B^*$ be the dual norm in $\mathcal{V}^*$, then the unit ball $\mathcal{V}_1^*$ for $\|\cdot\|_B^*$ is given by

$$\mathcal{V}_1^* = \mathcal{V}_1^\circ = (co(B \cup -B))^\circ = (B \cup -B)^\circ = B^\circ \cap (-B)^\circ,$$

where $A^\circ := \{f \in \mathcal{V}^*,\ \langle f, a\rangle \leq 1, \forall a \in A\}$ is the polar of $A \subset \mathcal{V}$, see Ref. 18. We have

$$\mathcal{V}_1^* = \{f \in \mathcal{V}^*, -1 \leq \langle f, b\rangle \leq 1, \forall b \in B\} = \{f \in \mathcal{V}^*, -e_B \leq_{Q^*} f \leq_{Q^*} e_B\},$$

where $e_B$ is the order unit. Hence the dual norm is given by

$$\|f\|_B^* = \inf\{\lambda > 0, -\lambda e_B \leq_{Q^*} f \leq_{Q^*} \lambda e_B\} =: \|f\|_{e_B}.$$

In general, if $e$ is an order unit, then $\|\cdot\|_e$ defines a norm called the order unit norm in $\mathcal{V}^*$.

Since $\|\cdot\|_B$ is the dual norm for $\|\cdot\|_{e_B}$, we get for $v \in \mathcal{V}$,

$$\|v\|_B = \|v\|_{e_B}^* = \sup_{-e_B \leq_{Q^*} f \leq_{Q^*} e_B} \langle f, v\rangle = 2 \sup_{f \in Q^*, f \leq_{Q^*} e_B} \langle f, v\rangle - \langle e_B, v\rangle, \tag{4}$$

where the last equality follows by replacing $f$ by $\frac{1}{2}(f + e_B)$.

*Example 1.* Let $\mathcal{V} = B_h(\mathcal{H})$ and $Q = B(\mathcal{H})^+$. We identify $\mathcal{V}^*$ with $\mathcal{V}$, with duality $\langle a, b\rangle = \mathrm{Tr}\, ab$, then $Q$ is a self-dual proper cone and $B = \mathfrak{S}(\mathcal{H})$ is a base of $Q$, with $e_B = I$. The order unit norm $\|\cdot\|_I$ is the operator norm $\|\cdot\|$ in $B(\mathcal{H})$ and its dual $\|\cdot\|_B$ is the trace norm $\|\cdot\|_1$.

We will finish this section by showing that the base norm is naturally related to a distinguishability measure for elements of the base. By analogy with the set of quantum states, let us define a test on a base $B$ as an affine map $\mathbf{t} : B \to [0, 1]$. It is easy to see that there is a one-to-one correspondence between tests on $B$ and elements $f \le_{Q^*} e_B$ in $Q^*$. Let $b_0, b_1$ be two elements of $B$ and let us interpret the value $\mathbf{t}(b) = \langle f, b\rangle$ as the probability of choosing $b_0$ if the "true value" is $b$. Then $\langle f, b_1\rangle$ and $1 - \langle f, b_0\rangle$ are probabilities of making an error. Let $\lambda \ge 0$, then we define the minimal average error probability as

$$\Pi^B_\lambda(b_0, b_1) := \min_{0\le_{Q^*} f \le_{Q^*} e_B} \lambda(1 - \langle f, b_0\rangle) + (1 - \lambda)\langle f, b_1\rangle.$$

We obtain by (4) that

$$\Pi^B_\lambda(b_0, b_1) = \lambda - \max_{0\le_{Q^*} f \le_{Q^*} e_B} \langle f, \lambda b_0 - (1 - \lambda)b_1\rangle$$

$$= \frac{1}{2}(1 - \|\lambda b_0 - (1 - \lambda)b_1\|_B).$$

## II. BASE NORMS ON SUBSPACES OF $B_h(\mathcal{H})$

We now put $\mathcal{V} = B_h(\mathcal{H})$, with the self-dual proper cone $B(\mathcal{H})^+$ as in Example 1. We will describe all possible bases of this cone.

It is clear that $int(B(\mathcal{H})^+) = B(\mathcal{H})^{++}$, hence the strictly positive elements are the order units in $B_h(\mathcal{H})$. By Corollary 1, there is a one-to-one correspondence between strictly positive elements and bases of $B(\mathcal{H})^+$, given by

$$B(\mathcal{H})^{++} \ni b \leftrightarrow S_b := \{a \in B(\mathcal{H})^+, \mathrm{Tr}\, ab = 1\} = B(\mathcal{H})^+ \cap \mathcal{T}_b, \tag{5}$$

where $\mathcal{T}_b = \{x \in B_h(\mathcal{H}), \mathrm{Tr}\, xb = 1\}$. By (4) and Example 1, the corresponding base norm is

$$\|x\|_{S_b} = \sup_{-b\le a\le b} \mathrm{Tr}\, ax = \sup_{-I\le a\le I} \mathrm{Tr}\, ab^{1/2}xb^{1/2} = \|b^{1/2}xb^{1/2}\|_1 \tag{6}$$

and the dual order unit norm is

$$\|x\|_b = \inf\{\lambda > 0, -\lambda b \le x \le \lambda b\} = \|b^{-1/2}xb^{-1/2}\|. \tag{7}$$

If $b \in B(\mathcal{H})^+$, we define

$$\|b^{-1/2}xb^{-1/2}\| := \lim_{\varepsilon\to 0^+} \|(b + \varepsilon)^{-1/2}x(b + \varepsilon)^{-1/2}\|.$$

Note that the expression on the RHS is bounded for all $\varepsilon > 0$ if and only if $\mathrm{supp}\,(x) \le \mathrm{supp}\,(b)$ and in this case the norm on the LHS is defined by restriction to the support of $b$. Otherwise, the limit is infinite. Moreover, for $a, b \in B(\mathcal{H})^+$, we define

$$D_{max}(a\|b) := \log \inf\{\lambda > 0, a \le \lambda b\} = \inf\{\gamma > 0, a \le 2^\gamma b\}.$$

For a pair of states $\rho$ and $\sigma$, $D_{max}(\rho\|\sigma)$ is the max-relative entropy of $\rho$ and $\sigma$, introduced in Ref. 6. (Note that $D_{max}$ was denoted by $D_\infty$ in Ref. 17.) If $b \in B(\mathcal{H})^{++}$, then

$$D_{max}(a\|b) = \log(\|a\|_b).$$

In general, if $\mathrm{supp}\,(a) \le \mathrm{supp}\,(b)$, then we may restrict to the support of $b$ and with this restriction $D_{max}(a\|b) = \log(\|a\|_b)$, otherwise $D_{max}(a\|b) = \infty$.

### A. Sections of a base of $B(\mathcal{H})^+$

Let $J \subset B_h(\mathcal{H})$ be a subspace and let $Q = J \cap B(\mathcal{H})^+$ be the convex cone of positive elements in $J$. It is obvious that $Q$ is closed and pointed. We will suppose that $J$ is positively generated, then $J$

$= Q - Q$ and $Q$ is a proper cone in $J$. Let $b \in Q$ be such that $\mathrm{supp}\, a \leq \mathrm{supp}\, b =: p$ for all $a \in Q$, then $J \subseteq B_h(p\mathcal{H})$ and by restricting to $B_h(p\mathcal{H})$, we may suppose that $b$ is strictly positive. Conversely, if $J$ contains a strictly positive element, then $J$ is positively generated.

Let $J^\perp = \{y \in B_h(\mathcal{H}), \mathrm{Tr}\, xy = 0,\ \ x \in J\}$, let $B_h(\mathcal{H})|_{J^\perp}$ be the quotient space and let $\pi : B_h(\mathcal{H}) \to B_h(\mathcal{H})|_{J^\perp}$ be the quotient map $a \mapsto a + J^\perp$. We may identify the dual space $J^*$ with $B_h(\mathcal{H})|_{J^\perp}$, with duality

$$\langle x, \pi(a)\rangle = \mathrm{Tr}\, xa, \qquad x \in J,\ a \in B_h(\mathcal{H}).$$

It was shown in Theorem 2 of Ref. 13 that the dual cone of $Q$ is $Q^* = \pi(B(\mathcal{H})^+)$, moreover, since $\pi$ is a linear map, we have $int(Q^*) = int(\pi(B(\mathcal{H})^+)) = \pi(B(\mathcal{H})^{++})$ by Theorem 6.6 of Ref. 18. In other words, any element $f \in Q^*$ has the form

$$f(x) = \mathrm{Tr}\, ax, \qquad x \in J,$$

for some (in general non-unique) element $a \in B(\mathcal{H})^+$ and $f$ is an order unit in $J^*$ if and only if $a$ may be chosen strictly positive. Now we can use Corollary 1 to describe all bases of $Q$.

*Lemma 2. A subset $B \subset Q$ is a base of $Q$ if and only if $B = J \cap S_{\tilde{b}}$, where $\tilde{b} \in B(\mathcal{H}^{++})$. In this case, $\pi(\tilde{b}) = e_B$.*

*Proof.* Let $B$ be a base of $Q$. Since $e_B \in int(Q^*)$, there is some $\tilde{b} \in B(\mathcal{H})^{++}$ such that $e_B = \pi(\tilde{b})$ and

$$B = \{q \in Q, \mathrm{Tr}\, q\tilde{b} = \langle e_B, q\rangle = 1\rangle\} = Q \cap \mathcal{T}_{\tilde{b}} = J \cap S_{\tilde{b}}$$

(see (5)). Conversely, it is quite clear that $B = J \cap S_{\tilde{b}}$ is a base of $Q$ and $e_B = \pi(\tilde{b})$. □

A set of the form $B = L \cap S_{\tilde{b}}$ where $\tilde{b} \in B(\mathcal{H})^{++}$ and $L \subseteq B_h(\mathcal{H})$ is a subspace will be called a section of a base of $B(\mathcal{H})^+$, or simply a section. Let span($B$) be the real linear span of $B$, then

$$B \subseteq \mathrm{span}(B) \cap S_{\tilde{b}} \subseteq L \cap S_{\tilde{b}} = B,$$

so that $B = \mathrm{span}(B) \cap S_{\tilde{b}}$ and $B$ is a base of $\mathrm{span}(B) \cap B(\mathcal{H})^+$. If moreover $B$ contains a positive definite element, we say that $B$ is a faithful section. In this case, we have $B \cap B(\mathcal{H})^{++} = ri(B)$, where $ri(B)$ denotes the relative interior of $B$, Section 6 of Ref. 18. Indeed, since $B = L_{\tilde{b}} \cap B(\mathcal{H})^+$, where $L_{\tilde{b}} =: L \cap \mathcal{T}_{\tilde{b}}$ is an affine subspace containing an interior point of $B(\mathcal{H})^+$, we have by Corollary 6.5.1 of Ref. 18 that

$$ri(B) = ri(L_{\tilde{b}} \cap B(\mathcal{H})^+) = L_{\tilde{b}} \cap B(\mathcal{H})^{++} = B \cap B(\mathcal{H})^{++}.$$

For example, note that if $B = \{b\}$ for some $b \in B(\mathcal{H})^+$, then $B$ is a section and $B$ is faithful if and only if $b$ is strictly positive. If a section $B$ is not faithful, then there is some element $b \in B$ such that $p = \mathrm{supp}\,(b)$ and $B \subset B(p\mathcal{H})$. Then $B$ is a faithful section of a base of $B(p\mathcal{H})^+$, in this case, $ri(B) = B \cap ri(B(p\mathcal{H})^+)$. From now on, we will suppose that $B$ is a faithful section of a base of $B(\mathcal{H})^+$.

Note that in Lemma 2, the correspondence between the base $B$ and the element $\tilde{b}$ such that $B = \mathrm{span}(B) \cap S_{\tilde{b}}$ is not one-to-one, since the order unit $e_B = \pi(\tilde{b})$ may contain more different strictly positive elements. We will now look at the set of all such elements. Let

$$\tilde{B} := \{\tilde{b} \in B(\mathcal{H})^+, \mathrm{Tr}\, b\tilde{b} = 1, \forall b \in B\}.$$

Then

$$\tilde{B} = \pi^{-1}(e_B) \cap B(\mathcal{H})^+ = (\tilde{b} + B^\perp) \cap B(\mathcal{H})^+, \tag{8}$$

where $\tilde{b}$ is any element in $\tilde{B}$. Note that $\tilde{B}$ always contains a strictly positive element. Since by (8) $\tilde{B}$ is an intersection of $B(\mathcal{H})^+$ by an affine subspace, we have

$$\{\tilde{b} \in B(\mathcal{H})^{++},\ B = \mathrm{span}(B) \cap S_{\tilde{b}}\} = \tilde{B} \cap B(\mathcal{H})^{++} = ri(\tilde{B}).$$

*Lemma 3.*

*(i)* $\tilde{B}$ *is a faithful section of a base of* $B(\mathcal{H})^+$.
*(ii)* $\tilde{\tilde{B}} = B$.
*(iii)* $B = \bigcap_{\tilde{b}' \in ri(\tilde{B})} S_{\tilde{b}'}$.

*Proof.* (i) Let $b \in ri(B)$. Since $\tilde{B}$ is convex, any element $y \in \text{span}(\tilde{B})$ has the form $y = \lambda\tilde{b}_1 - \mu\tilde{b}_2$, with $\tilde{b}_1, \tilde{b}_2 \in \tilde{B}$ and $\lambda,\ \mu \geq 0$. Hence by (8), $y = (\lambda - \mu)\tilde{b} + z$ for some $z \in B^{\perp}$. If $y$ is also in $S_b$, we must have $1 = \text{Tr}\, yb = \lambda - \mu$, so that $y \in (\tilde{b} + B^{\perp}) \cap B(\mathcal{H})^+ = \tilde{B}$. It follows that $\tilde{B} = \text{span}(\tilde{B}) \cap S_b$.

(ii) It is clear that $B \subseteq \tilde{\tilde{B}}$ and $\tilde{\tilde{B}} = (b + \tilde{B}^{\perp}) \cap B(\mathcal{H})^+$. Let $\tilde{b} \in ri(\tilde{B})$, then $\tilde{B} = (\tilde{b} + B^{\perp}) \cap B(\mathcal{H})^+$. Since $\tilde{b} \in B(\mathcal{H})^{++}$, for each $z \in B^{\perp}$ there is some $t > 0$ such that $\tilde{b} + tz \in \tilde{B}$ and this implies that $\tilde{B}^{\perp} \subseteq (B^{\perp})^{\perp} = \text{span}(B)$, hence also $\tilde{\tilde{B}} \subseteq \text{span}(B)$. It follows that $\text{span}(B) = \text{span}(\tilde{\tilde{B}})$, so that $B$ and $\tilde{\tilde{B}}$ are two bases of the same cone. This implies (ii).

(iii) Obviously $B \subseteq \bigcap_{\tilde{b}' \in ri(\tilde{B})} S_{\tilde{b}'}$. If $a \in \bigcap_{\tilde{b}' \in ri(\tilde{B})} S_{\tilde{b}'}$, then $a$ is a positive element such that $\text{Tr}\, a\tilde{b}' = 1$ for all $\tilde{b}' \in cl(ri(\tilde{B})) = \tilde{B}$, hence $a \in \tilde{\tilde{B}} = B$. □

We call $\tilde{B}$ the dual section of $B$. The section $B$ defines a base norm $\|\cdot\|_B$ in span($B$). Next we show that this norm can be naturally extended to all $B_h(\mathcal{H})$. For this, let us define

$$\mathcal{O}_B := \{x \in B_h(\mathcal{H}), x = x_1 - x_2,\ x_1, x_2 \in B(\mathcal{H})^+, x_1 + x_2 \in B\}. \tag{9}$$

For $b \in B(\mathcal{H})^+$, we define $\mathcal{O}_b := \mathcal{O}_{\{b\}}$.

*Lemma 4. We have*

*(i)* $\mathcal{O}_B = \{x \in B_h(\mathcal{H}), \exists b' \in B,\ -b' \leq x \leq b'\} = \bigcup_{b' \in B} \mathcal{O}_{b'}$.
*(ii)* *The unit ball of the base norm* $\|\cdot\|_B$ *is* $\mathcal{O}_B \cap \text{span}(B)$.

*Proof.* (i) Let $x = x_1 - x_2$ with $x_1 + x_2 = b' \in B$, then $-b' = -(x_1 + x_2) \leq x \leq x_1 + x_2 = b'$. Conversely, let $-b' \leq x \leq b'$ for some $b' \in B$. Put $x_{\pm} = 1/2(b' \pm x)$, then $x_{\pm} \in B(\mathcal{H})^+$, $x_+ - x_- = x$, and $x_+ + x_- = b' \in B$.

(ii) By definition, the unit ball of $\|\cdot\|_B$ is the set of elements of the form $x = \lambda b_1 - (1 - \lambda)b_2$, $b_1, b_2 \in B$, $0 \leq \lambda \leq 1$. Then clearly $x \in \mathcal{O}_B$, by putting $x_1 = \lambda b_1$ and $x_2 = (1 - \lambda)b_2$. Conversely, let $x \in \text{span}(B)$ be such that $-b' \leq x \leq b'$ for some $b' \in B$, then $x_{\pm} = 1/2(b' \pm x)$ are positive elements in span($B$) and we have $x_{\pm} = \lambda_{\pm} b_{\pm}$, for $\lambda_{\pm} \geq 0$, $b_{\pm} \in B$. By applying the order unit $e_B$ to the equality $b' = x_+ + x_-$, we see that we must have $\lambda_+ + \lambda_- = 1$, so that $\|x\|_B \leq 1$. □

**Theorem 1.** *Let B be a faithful section and let* $\tilde{B}$ *be the dual section. Then* $\mathcal{O}_B$ *is the unit ball of a norm in* $B_h(\mathcal{H})$. *The unit ball of the dual norm is* $\mathcal{O}_{\tilde{B}}$.

We will denote this norm by $\|\cdot\|_B$, note that Lemma 4 (ii) justifies this notation.

*Proof.* It is clear that $\mathcal{O}_B$ is convex and symmetric, that is, $-\mathcal{O}_B \subseteq \mathcal{O}_B$. Since $B$ is compact, $\mathcal{O}_B$ is closed. If $x \in \mathcal{O}_B$, then $x = x_1 - x_2$ with $x_1, x_2 \geq 0$, $x_1 + x_2 \in B$ and by (6),

$$\|x\|_{S_{\tilde{b}}} \leq \|x_1\|_{S_{\tilde{b}}} + \|x_2\|_{S_{\tilde{b}}} = \text{Tr}\,(x_1 + x_2)\tilde{b} = 1,$$

for any $\tilde{b} \in ri(\tilde{B})$, hence $\mathcal{O}_B$ is bounded. Moreover, since $b \in ri(B)$ is an order unit, for every $x \in B_h(\mathcal{H})$ there is some $t > 0$ such that $-tb \leq x \leq tb$, so that $x \in t\mathcal{O}_B$ (see Lemma 4 (i)). This means that $\mathcal{O}_B$ is absorbing . These facts imply that $\mathcal{O}_B$ is the unit ball of a norm.

To show duality of the norms $\|\cdot\|_B$ and $\|\cdot\|_{\tilde{B}}$, let $\mathcal{H}_2 = \mathcal{H} \oplus \mathcal{H}$ and let $\Phi : B_h(\mathcal{H}_2) \to B_h(\mathcal{H})$ be the map defined by $\Phi(a \oplus b) = a + b$. Let $J_2 = \Phi^{-1}(\text{span}(B))$, then $J_2$ is a subspace in $B_h(\mathcal{H}_2)$ and

$$J_2^{\perp} = \Phi^*(B^{\perp}) = \{x \oplus x,\ x \in B^{\perp}\},$$

see Ref. 13. Let $\pi_2 : B(\mathcal{H}_2) \to J_2^* = B(\mathcal{H}_2)|_{J_2^{\perp}}$ be the quotient map.

Let $\tilde{b} \in ri(\tilde{B})$ and put $B_2 = J_2 \cap S_{\tilde{b}\oplus\tilde{b}}$. Then $B_2$ is a base of $Q_2 = J_2 \cap B(\mathcal{H}_2)^+$ and it is clear that for $w_1, w_2 \in B(\mathcal{H})^+$, $w_1 \oplus w_2 \in B_2$ if and only if $w_1 + w_2 \in B$. Let now $a \in B_h(\mathcal{H})$, then $a \in \mathcal{O}_B^\circ$ if and only if $\mathrm{Tr}\,(a \oplus -a)w \leq 1$ for all $w \in B_2$. Equivalently,

$$\pi_2(a \oplus -a) \leq_{Q_2^*} e_{B_2} = \pi_2(\tilde{b} \oplus \tilde{b}),$$

that is, there is some $v \in J_2^\perp$ such that $a \oplus -a \leq \tilde{b} \oplus \tilde{b} + v$. Since $v = x \oplus x$, $x \in B^\perp$, we obtain $\pm a \leq \tilde{b} + x$. Note that we must have $\tilde{b} + x \geq 0$: if $c$ is any element in $B(\mathcal{H})^+$, then we have $\pm\,\mathrm{Tr}\,ca \leq \mathrm{Tr}\,c(b + x)$, so that $\mathrm{Tr}\,c(b + x)$ cannot be negative. Hence $\pm a \leq \tilde{b} + x \in \tilde{B}$, so that $a \in \mathcal{O}_{\tilde{B}}$, by Lemma 4 (i). This shows that $\mathcal{O}_B^\circ \subseteq \mathcal{O}_{\tilde{B}}$. Conversely, it is easy to see that if $-b \leq x \leq b$ and $-\tilde{b} \leq y \leq \tilde{b}$ for $b \in B$, $\tilde{b} \in \tilde{B}$, then $\mathrm{Tr}\,xy \leq \mathrm{Tr}\,b\tilde{b} = 1$, this implies the opposite inclusion. □

*Corollary 2. Let $x \in B_h(\mathcal{H})$. Then*

*(i)* $\mathcal{O}_B = \bigcap_{\tilde{b}\in ri(\tilde{B})} \mathcal{O}_{S_{\tilde{b}}}$,
*(ii)* $\|x\|_B = \sup_{\tilde{b}\in ri(\tilde{B})} \|x\|_{S_{\tilde{b}}} = \sup_{\tilde{b}\in\tilde{B}} \|\tilde{b}^{1/2} x \tilde{b}^{1/2}\|_1$,
*(iii)* $\|x\|_B = \in f_{b\,\in\, ri(B)} \|x\|_b = \in f_{b\,\in\, B} \|b^{-1/2} x b^{-1/2}\|$.

*Proof.* (i) It is easy to see from Lemma 4 that

$$\mathcal{O}_B = \bigcup_{b\in B} \mathcal{O}_b = cl(\bigcup_{b\in ri(B)} \mathcal{O}_b). \tag{10}$$

Indeed, let $x \in B_h(\mathcal{H})$ be such that $-b \leq x \leq b$ for some $b \in B$ and let $b' \in ri(B)$, then $b_\epsilon := \epsilon b' + (1 - \epsilon)b \in ri(B)$ for all $0 < \epsilon < 1$. Let $x' \in \mathcal{O}_{b'}$ be any element, then $x_\epsilon := \epsilon x' + (1 - \epsilon)x \in \mathcal{O}_{b_\epsilon}$ and $x = \lim_{\epsilon\to 0^+} x_\epsilon \in cl(\bigcup_{b\in ri(B)} \mathcal{O}_b)$.

Since $A^\circ = (cl(conv(A)))^\circ$ for any subset $A \in B_h(\mathcal{H})$ containing 0, we obtain by Theorem 1 that

$$\mathcal{O}_B = \mathcal{O}_{\tilde{B}}^\circ = (\bigcup_{\tilde{b}\in ri(\tilde{B})} \mathcal{O}_{\tilde{b}})^\circ = \bigcap_{\tilde{b}\in ri(\tilde{B})} \mathcal{O}_{\tilde{b}}^\circ = \bigcap_{\tilde{b}\in ri(\tilde{B})} \mathcal{O}_{S_{\tilde{b}}}.$$

(ii) Since $\mathcal{O}_B$ is the unit ball of $\|\cdot\|_B$, we get from (i)

$$\begin{aligned}\|x\|_B &= \inf\{\lambda > 0, x \in \lambda\mathcal{O}_B\} = \inf\{\lambda > 0, x \in \lambda\mathcal{O}_{S_{\tilde{b}}}, \forall\tilde{b} \in ri(\tilde{B})\}\\ &= \inf\{\lambda > 0, \lambda \geq \|x\|_{S_{\tilde{b}}}, \forall\tilde{b} \in ri(\tilde{B})\} = \sup_{\tilde{b}\in ri(\tilde{B})} \|x\|_{S_{\tilde{b}}} = \sup_{\tilde{b}\in\tilde{B}} \|\tilde{b}^{1/2} x \tilde{b}^{1/2}\|_1,\end{aligned}$$

the last equality follows from (6) and continuity of the norm $\|\cdot\|_1$.

(iii) On the other hand, we get from Lemma 4 and (10)

$$\begin{aligned}\|x\|_B &= \inf\{\lambda > 0, x \in \lambda\mathcal{O}_B\} = \inf\{\lambda > 0, x \in \lambda \cup_{b\in ri(B)} \mathcal{O}_b\}\\ &= \inf_{b\in ri(B)} \inf\{\lambda > 0, x \in \lambda\mathcal{O}_b\} = \inf_{b\in ri(B)} \|x\|_b = \inf_{b\in B} \|b^{-1/2} x b^{-1/2}\|,\end{aligned}$$

where the last equality follows by (7). □

*Corollary 3. For $a \in B(\mathcal{H})^+$, we have*

$$\|a\|_B = \sup_{\tilde{b}\in\tilde{B}} \mathrm{Tr}\,a\tilde{b} = \inf_{b\in B} 2^{D_{max}(a\|b)}.$$

*Proof.* We have

$$\|a\|_B = \sup_{x\in\mathcal{O}_{\tilde{B}}} \mathrm{Tr}\,ax.$$

Let $x \in \mathcal{O}_{\tilde{B}}$, then $x = x_1 - x_2$, $x_1, x_2 \in B(\mathcal{H})^+$ and $x_1 + x_2 =: \tilde{b}_x \in \tilde{B}$, so that

$$\mathrm{Tr}\,ax \leq \mathrm{Tr}\,ax_1 \leq \mathrm{Tr}\,a\tilde{b}_x \leq \sup_{\tilde{b}\in\tilde{B}} \mathrm{Tr}\,a\tilde{b} \leq \sup_{y\in\mathcal{O}_{\tilde{B}}} \mathrm{Tr}\,ay = \|a\|_B.$$

Hence $\|a\|_B = \sup_{\tilde{b}\in\tilde{B}} \mathrm{Tr}\, a\tilde{b}$. The second equality follows directly from Corollary 2 (iii) and the definition of $D_{max}$. □

We can also characterize the maximizer resp. minimizer in Corollary 3.

*Corollary 4. Let $a \in B(\mathcal{H})^+$.*

*(i) Let $\tilde{b}_0 \in \tilde{B}$, then $\|a\|_B = \mathrm{Tr}\, a\tilde{b}_0$ if and only if there exists some $q \in \mathrm{span}(B)$, such that $a \leq q$ and $(q - a)\tilde{b}_0 = 0$. In this case, $q = \|a\|_B b_0$, $b_0 \in B$, and $\|a\|_B = 2^{D_{max}(a\|b_0)}$.*

*(ii) Let $b_0 \in B$, then $\|a\|_B = 2^{D_{max}(a\|b_0)}$ if and only if there exists some $t > 0$ and $\tilde{b}_0 \in \tilde{B}$, such that $a \leq tb_0$ and $(tb_0 - a)\tilde{b}_0 = 0$. In this case, $t = \|a\|_B = \mathrm{Tr}\, a\tilde{b}_0$.*

*Proof.* (i) Let $\tilde{b}_0 \in \tilde{B}$ be such that $\|a\|_B = \mathrm{Tr}\, a\tilde{b}_0$. Let $b_0 \in B$ be such that $\|a\|_B = 2^{D_{max}(a\|b_0)}$, in particular, $a \leq \|a\|_B b_0$. Put $q = \|a\|_B b_0$, then $q - a \geq 0$ and $\mathrm{Tr}\,(q - a)\tilde{b}_0 = 0$. Since also $\tilde{b}_0 \geq 0$, it follows that $(q - a)\tilde{b}_0 = 0$.

Conversely, suppose $q \in \mathrm{span}(B)$ satisfies $a \leq q$ and $(q - a)\tilde{b}_0 = 0$. Then $q = sb_0$ for some $b_0 \in B$, $s \geq 0$. Since $a \leq sb_0$, we have

$$\|a\|_B \leq s = \mathrm{Tr}\, a\tilde{b}_0 \leq \|a\|_B,$$

so that $\mathrm{Tr}\, a\tilde{b}_0 = \|a\|_B = s = 2^{D_{max}(a\|b_0)}$.

(ii) is proved similarly. □

## III. GENERALIZED CHANNELS

Let $B$ be a section of a base of $B(\mathcal{H})^+$. A generalized channel with respect to $B$ (or a $B$-channel) is a completely positive map $\Phi : B(\mathcal{H}) \to B(\mathcal{K})$ such that $\Phi(B) \subseteq \mathfrak{S}(\mathcal{K})$. Let $X_\Phi$ be the Choi matrix of $\Phi$, then $\Phi$ is a generalized channel with respect to $B$ if and only if $X_\Phi \geq 0$ and

$$1 = \mathrm{Tr}\,\Phi(b) = \mathrm{Tr}\,\mathrm{Tr}_{\mathcal{H}}[(I \otimes b^\mathsf{T})X_\Phi] = \mathrm{Tr}\,(I \otimes b^\mathsf{T})X_\Phi = \mathrm{Tr}\, b^\mathsf{T}\mathrm{Tr}_{\mathcal{K}} X_\Phi,$$

for all $b \in B$. Let $\mathcal{C}_B(\mathcal{H}, \mathcal{K})$ denote the set of Choi matrices of all generalized channels with respect to $B$, then

$$\mathcal{C}_B(\mathcal{H}, \mathcal{K}) = \{X \in B(\mathcal{K} \otimes \mathcal{H})^+, \mathrm{Tr}_{\mathcal{K}} X \in \tilde{B}^\mathsf{T}\}.$$

Let us remark that if $B$ is a section, then $B^\mathsf{T} := \{b^\mathsf{T}, b \in B\}$ is a section as well, here $b^\mathsf{T}$ denotes the transpose of $b$. Moreover, $\widetilde{B^\mathsf{T}} = \tilde{B}^\mathsf{T}$. Note also that we have

$$\mathcal{C}_B(\mathcal{H}, \mathbb{C}) = \tilde{B}^\mathsf{T}, \tag{11}$$

so that, in particular, $\mathcal{C}_B(\mathcal{H}, \mathbb{C})$ is a section.

*Proposition 1. Let B be a faithful section of a base of $B(\mathcal{H})^+$. Then $\mathcal{C}_B(\mathcal{H}, \mathcal{K})$ is a faithful section of a base of $B(\mathcal{K} \otimes \mathcal{H})^+$ and $\widetilde{\mathcal{C}_B(\mathcal{H}, \mathcal{K})} = \{I_{\mathcal{K}} \otimes b^\mathsf{T},\ b \in B\}$.*

*Proof.* It is easy to see that $I_{\mathcal{K}} \otimes B^\mathsf{T} = \{I_{\mathcal{K}} \otimes b^\mathsf{T}, b \in B\}$ is a faithful section of a base of $B(\mathcal{K} \otimes \mathcal{H})^+$ and

$$\mathcal{C}_B(\mathcal{H}, \mathcal{K}) = \{X \in B(\mathcal{K} \otimes \mathcal{H})^+, \mathrm{Tr}\, X(I \otimes b^\mathsf{T}) = 1, \forall b \in B\} = \widetilde{I_{\mathcal{K}} \otimes B^\mathsf{T}}.$$

The proof now follows by Lemma 3 (i) and (ii). □

Let now $X \in B_h(\mathcal{K} \otimes \mathcal{H})$ and let $\Phi : B(\mathcal{H}) \to B(\mathcal{K})$ be the corresponding Hermitian map. By Corollary 2 and Proposition 1,

$$\|X\|_{\mathcal{C}_B(\mathcal{H},\mathcal{K})} = \sup_{b\in B} \|(I \otimes (b^\mathsf{T})^{1/2})X(I \otimes (b^\mathsf{T})^{1/2})\|_1$$

and we have

$$(I \otimes (b^\mathsf{T})^{1/2})X(I \otimes (b^\mathsf{T})^{1/2}) = (\Phi \otimes id_{\mathcal{H}})(\sigma_b),$$

where $\sigma_b = |\psi_b\rangle\langle\psi_b|$, with

$$|\psi_b\rangle = \sum_i |i\rangle \otimes (b^{\mathsf{T}})^{1/2}|i\rangle = \sum_i b^{1/2}|i\rangle \otimes |i\rangle \in \mathcal{H} \otimes \mathcal{H}.$$

Hence $\sigma_b \in B(\mathcal{H}\otimes\mathcal{H})^+$ and $\mathrm{Tr}_1\sigma_b = b^{\mathsf{T}} \in B^{\mathsf{T}}$, so that $\sigma_b \in \mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{H})$. Conversely, if $\sigma = |\varphi\rangle\langle\varphi| \in \mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L})$ for some Hilbert space $\mathcal{L}$, then there is some linear map $R : \mathcal{H} \to \mathcal{L}$ satisfying $R^*R = b \in B$ and such that $|\varphi\rangle = \sum_i R|i\rangle \otimes |i\rangle$. Let $U : \mathcal{H} \to \mathcal{L}$ be an isometry such that $R = Ub^{1/2}$, then

$$|\varphi\rangle = \sum_i R|i\rangle \otimes |i\rangle = \sum_i Ub^{1/2}|i\rangle \otimes |i\rangle = (U \otimes I)|\psi_b\rangle.$$

**Theorem 2.** *Let $X \in B_h(\mathcal{K}\otimes\mathcal{H})$ and let $\Phi$ be the corresponding Hermitian map $B(\mathcal{H}) \to B(\mathcal{K})$. Let $\mathcal{L}$ be any Hilbert space with* $\dim(\mathcal{L}) = \dim(\mathcal{H})$. *Then*

$$\|X\|_{\mathcal{C}_B(\mathcal{H},\mathcal{K})} = \sup_{\dim(\mathcal{L}')<\infty}\ \sup_{\sigma\in\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L}')} \|(\Phi\otimes id_{\mathcal{L}'})(\sigma)\|_1$$
$$= \sup_{\sigma\in\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L})} \|(\Phi\otimes id_{\mathcal{L}})(\sigma)\|_1,$$

*and the dual norm is $\|X\|^*_{\mathcal{C}_B(\mathcal{H},\mathcal{K})} = \|X\|_{I_\mathcal{K}\otimes B^{\mathsf{T}}}$. Moreover, if $X \geq 0$ then*

$$\|X\|_{\mathcal{C}_B(\mathcal{H},\mathcal{K})} = \sup_{b\in B} \mathrm{Tr}\,\Phi(b) = \inf_{Y\in\mathcal{C}_B(\mathcal{H},\mathcal{K})} 2^{D_{max}(X\|Y)}$$

*and*

$$\|X\|_{I\otimes B^{\mathsf{T}}} = \inf_{b\in B} 2^{D_{max}(X\|I\otimes b^{\mathsf{T}})} = \sup_{Y\in\mathcal{C}_B(\mathcal{H},\mathcal{K})} \mathrm{Tr}\,XY = \sup_S \langle\psi|X_{S^*\circ\Phi}|\psi\rangle,$$

*where the last supremum is taken over the set of all B-channels $B(\mathcal{H}) \to B(\mathcal{K})$.*

*Proof.* From what was said above, it is easy to see that

$$\|X\|_{\mathcal{C}_B(\mathcal{H},\mathcal{K})} = \sup_{|\varphi\rangle\langle\varphi|\in\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L})} \|(\Phi\otimes id_{\mathcal{L}})(|\varphi\rangle\langle\varphi|)\|_1,$$

with $\dim(\mathcal{L}) = \dim(\mathcal{H})$. We will show that

$$\sup_{|\varphi\rangle\langle\varphi|\in\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L}')} \|(\Phi\otimes id_{\mathcal{L}'})(|\varphi\rangle\langle\varphi|)\|_1 \leq \sup_{|\varphi\rangle\langle\varphi|\in\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L})} \|(\Phi\otimes id_{\mathcal{L}})(|\varphi\rangle\langle\varphi|)\|_1,$$

whenever $\dim(\mathcal{L}') \geq \dim(\mathcal{L})$. The proof is almost the same as the proof of Theorem 5 of Ref. 19, we include it here for completeness.

So let $\dim(\mathcal{L}') \geq \dim(\mathcal{L}) = \dim(\mathcal{H})$, then there is some $\varphi_0 \in \mathcal{H}\otimes\mathcal{L}'$, with $|\varphi_0\rangle\langle\varphi_0| \in \mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L}')$ such that

$$\sup_{|\varphi\rangle\langle\varphi|\in\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L}')} \|\Phi\otimes id_{\mathcal{L}'}(|\varphi\rangle\langle\varphi|)\|_1 = \|\Phi\otimes id_{\mathcal{L}'}(|\varphi_0\rangle\langle\varphi_0|)\|_1.$$

Let $|\varphi_0\rangle = \sum_{i=1}^m s_i|\varphi_i\rangle \otimes |\xi_i\rangle$ be the Schmidt decomposition of $\varphi_0$, with $\{|\varphi_i\rangle\}$ and $\{|\xi_i\rangle\}$ orthonormal sets in $\mathcal{H}$ resp. $\mathcal{L}'$ and $m = \dim(\mathcal{H})$. Then $|\varphi_0\rangle\langle\varphi_0| = \sum_{i,j}|\varphi_i\rangle\langle\varphi_j| \otimes |\xi_i\rangle\langle\xi_j|$ and

$$(\mathrm{Tr}_{\mathcal{L}'}|\varphi_0\rangle\langle\varphi_0|)^{\mathsf{T}} = (\sum_i s_i|\varphi_i\rangle\langle\varphi_i|)^{\mathsf{T}} \in \tilde{\tilde{B}} = B.$$

Let $\{|e_i\rangle,\ i = 1, \ldots, m\}$ be an ONB in $\mathcal{L}$. Define the linear map $U : \mathcal{L}' \to \mathcal{L}$ by $U = \sum_{i=1}^m |e_i\rangle\langle\xi_i|$, then $U^*U = \sum_i|\xi_i\rangle\langle\xi_i|$ is the projection in $\mathcal{L}'$ onto the subspace spanned by the vectors $|\xi_i\rangle$, $i = 1, \ldots, m$, and $(I\otimes U^*U)|\varphi_0\rangle = |\varphi_0\rangle$. Put $\varphi_U := (I\otimes U)|\varphi_0\rangle = \sum_i|\varphi_i\rangle \otimes |e_i\rangle$, then it is easy to see that $|\varphi_U\rangle\langle\varphi_U| \in \mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L})$. Now we have

$$\begin{aligned}\sup_{|\varphi\rangle\langle\varphi|\in\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L})} \|\Phi\otimes id_{\mathcal{L}}(|\varphi\rangle\langle\varphi|)\|_1 &\geq \|\Phi\otimes id_{\mathcal{L}}(|\varphi_U\rangle\langle\varphi_U|)\|_1\\ &\geq \|(I\otimes U^*)(\Phi\otimes id_{\mathcal{L}})(|\varphi_U\rangle\langle\varphi_U|)(I\otimes U)\|_1\\ &= \|\Phi\otimes id_{\mathcal{L}'}((I\otimes U^*)|\varphi_U\rangle\langle\varphi_U|(I\otimes U))\|_1\\ &= \|\Phi\otimes id_{\mathcal{L}'}(|\varphi_0\rangle\langle\varphi_0|)\|_1\\ &= \sup_{|\varphi\rangle\langle\varphi|\in\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L}')} \|\Phi\otimes id_{\mathcal{L}'}(|\varphi\rangle\langle\varphi|)\|_1.\end{aligned}$$

Next, let $Y$ be any element in $\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L}')$, then the corresponding map $\xi : B(\mathcal{H}) \to B(\mathcal{L}')$ has the form

$$\xi(a) = \sum_{i=1}^{N} V_i a V_i^*, \qquad a \in B(\mathcal{H}),$$

where $V_i : \mathcal{H} \to \mathcal{L}'$ are linear maps such that $\sum_i V_i^* V_i \in B$. Let $\mathcal{L}'_0$ be a Hilbert space with $\dim(\mathcal{L}'_0) = N$ and let $\{|f_j\rangle, j = 1, \ldots, N\}$ be an ONB in $\mathcal{L}'_0$. Define $V = \sum_{j=1}^N V_j \otimes |f_j\rangle$, then $V$ is a linear map $\mathcal{H} \to \mathcal{L}' \otimes \mathcal{L}'_0$ with $V^*V = \sum_i V_i^* V_i \in B$. Let $\mathcal{V}(a) = VaV^*$ and let $Z$ be the Choi matrix of $\mathcal{V}$, then $Z$ is a rank one element in $\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L}'\otimes\mathcal{L}'_0)$. Moreover, $\xi(a) = \mathrm{Tr}_{\mathcal{L}'_0} VaV^*$ and $Y = \mathrm{Tr}_{\mathcal{L}'_0} Z$. It follows that

$$\begin{aligned}\|(\Phi\otimes id_{\mathcal{L}'})(Y)\|_1 = \|(\Phi\otimes id_{\mathcal{L}'})(\mathrm{Tr}_{\mathcal{L}'_0} Z)\|_1 &= \|\mathrm{Tr}_{\mathcal{L}'_0}(\Phi\otimes id_{\mathcal{L}'\otimes\mathcal{L}'_0})(Z)\|_1\\ &\leq \|(\Phi\otimes id_{\mathcal{L}'\otimes\mathcal{L}'_0})(Z)\|_1 \leq \|X\|_{\mathcal{C}_B(\mathcal{H},\mathcal{K})}.\end{aligned}$$

We now have

$$\begin{aligned}\|X\|_{\mathcal{C}_B(\mathcal{H},\mathcal{K})} &= \sup_{|\varphi\rangle\langle\varphi|\in\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L})} \|(\Phi\otimes id_{\mathcal{L}})(|\varphi\rangle\langle\varphi|)\|_1 \leq \sup_{\sigma\in\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L})} \|(\Phi\otimes id_{\mathcal{L}})(\sigma)\|_1\\ &\leq \sup_{\dim(\mathcal{L}')<\infty}\ \sup_{\sigma\in\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{L}')} \|(\Phi\otimes id_{\mathcal{L}'})(\sigma)\|_1 \leq \|X\|_{\mathcal{C}_B(\mathcal{H},\mathcal{K})}.\end{aligned}$$

The expression for the dual norm follows by Proposition 1. Suppose now that $X \geq 0$, then by Corollary 3

$$\|X\|_{\mathcal{C}_B(\mathcal{H},\mathcal{K})} = \sup_{b\in B} \mathrm{Tr}\, X(I\otimes b^{\mathsf{T}}) = \inf_{Y\in\mathcal{C}_B(\mathcal{H},\mathcal{K})} 2^{D_{max}(X\|Y)},$$

$$\|X\|_{I\otimes B^{\mathsf{T}}} = \sup_{Y\in\mathcal{C}_B(\mathcal{H},\mathcal{K})} \mathrm{Tr}\, XY = \inf_{b\in B} 2^{D_{max}(X\|I\otimes b^{\mathsf{T}})}.$$

By (1), $\mathrm{Tr}\, X(I\otimes b^{\mathsf{T}}) = \mathrm{Tr}\,\mathrm{Tr}_{\mathcal{H}} X(I\otimes b^{\mathsf{T}}) = \mathrm{Tr}\,\Phi(b)$. Moreover, let $Y \in \mathcal{C}_B(\mathcal{H},\mathcal{K})$ and let $S$ be the corresponding $B$-channel, then

$$\mathrm{Tr}\, XY = \mathrm{Tr}\, X(S\otimes id)(\Psi) = \mathrm{Tr}\,(S^*\otimes id)(X)\Psi = \langle\psi, X_{S^*\circ\Phi}\psi\rangle.$$

□

### A. Channels

Let $B = \mathfrak{S}(\mathcal{H})$, then generalized channels are the usual channels. In this case, we denote $\mathcal{C}_B(\mathcal{H},\mathcal{K})$ by $\mathcal{C}(\mathcal{H},\mathcal{K})$. Note that $\tilde{B} = \{I\}$ and $\mathcal{C}_{\tilde{B}}(\mathcal{H},\mathcal{K}) = \mathfrak{S}(\mathcal{K}\otimes\mathcal{H})$.

By Proposition 1, $\mathcal{C}(\mathcal{H},\mathcal{K})$ is a faithful section of a base of $B(\mathcal{K}\otimes\mathcal{H})^+$ and

$$\widetilde{\mathcal{C}(\mathcal{H},\mathcal{K})} = \{I_{\mathcal{K}}\otimes\rho,\ \rho\in\mathfrak{S}(\mathcal{H})\}.$$

Furthermore, let $X \in B_h(\mathcal{K}\otimes\mathcal{H})$ and let $\Phi : B(\mathcal{H}) \to B(\mathcal{K})$ be the corresponding Hermitian map. Then by Theorem 2,

$$\|X\|_{\mathcal{C}(\mathcal{H},\mathcal{K})} = \sup_{\sigma\in\mathfrak{S}(\mathcal{H}\otimes\mathcal{L})} \|(\Phi\otimes id_{\mathcal{L}})(\sigma)\|_1 = \|\Phi\|_\diamond,$$

with $\dim(\mathcal{L}) = \dim(\mathcal{H})$. For the dual norm, we have

$$\|X\|_{I\otimes\mathfrak{S}(\mathcal{H})} = \inf_{\rho\in\mathfrak{S}(\mathcal{H})} \inf\{\lambda > 0,\ -\lambda(I\otimes\rho) \leq X \leq \lambda(I\otimes\rho)\}.$$

If $\sigma \in B(\mathcal{K}\otimes\mathcal{H})^+$, we obtain

$$\|\sigma\|_{I\otimes\mathfrak{S}(\mathcal{H})} = \inf_{\rho\in\mathfrak{S}(\mathcal{H})} 2^{D_{max}(\sigma\|I\otimes\rho)} = 2^{-H_{min}(\mathcal{K}|\mathcal{H})_\sigma},$$

where $H_{min}(\mathcal{K}|\mathcal{H})_\sigma$ is the conditional min-entropy, see Ref. 17.

### B. Quantum supermaps

Let $\mathcal{H}_0, \mathcal{H}_1, \ldots$ be a sequence of finite dimensional Hilbert spaces. For each $n \geq 1$, we define the sets $\mathcal{C}(\mathcal{H}_0, \ldots, \mathcal{H}_n)$ as follows: $\mathcal{C}(\mathcal{H}_0, \mathcal{H}_1)$ is, as before, the set of Choi matrices of channels $B(\mathcal{H}_0) \to B(\mathcal{H}_1)$. For $n > 1$, we define $\mathcal{C}(\mathcal{H}_0, \ldots, \mathcal{H}_n)$ as the set of Choi matrices of cp maps $B(\mathcal{H}_{n-1}\otimes\cdots\otimes\mathcal{H}_0) \to B(\mathcal{H}_n)$ that map $\mathcal{C}(\mathcal{H}_0, \ldots, \mathcal{H}_{n-1})$ into $\mathfrak{S}(\mathcal{H}_n)$. Such maps were called quantum supermaps in Ref. 13. (Note that this definition is slightly different from the notion of supermap introduced in Ref. 5, which is a cp map that maps Choi matrices of channels to Choi matrices of channels.) and it was proved that for $n = 2N - 1$ we get precisely the set of deterministic quantum $N$-combs for the sequence $\{\mathcal{H}_0, \ldots, \mathcal{H}_{2N-1}\}$, Theorem 7 of Ref. 13. If $n = 2N$, we get the set of $N + 1$-combs for $\{\mathbb{C}, \mathcal{H}_0, \ldots, \mathcal{H}_{2N}\}$.

Let us fix the sequence $\mathcal{H}_0, \mathcal{H}_1, \ldots$ and for this, put $\mathcal{C}_n = \mathcal{C}(\mathcal{H}_0, \ldots, \mathcal{H}_n)$. By using repeatedly Proposition 1, we see that $\mathcal{C}_n$ is a faithful section of a base of $B(\mathcal{H}_n\otimes\cdots\otimes\mathcal{H}_0)^+$ and

$$\mathcal{C}_{n+1} = \mathcal{C}_{\mathcal{C}_n}(\mathcal{H}_n\otimes\cdots\otimes\mathcal{H}_0, \mathcal{H}_{n+1}).$$

Moreover, by Proposition 1,

$$\widetilde{\mathcal{C}_n} = I_{\mathcal{H}_n}\otimes\mathcal{C}_{n-1} = \mathcal{C}(\mathcal{H}_0, \ldots, \mathcal{H}_n, \mathbb{C})$$

(note that $\mathcal{C}_{n-1}^{\mathsf{T}} = \mathcal{C}_{n-1}$, the last equality above follows from (11)). For $n = 2N - 1$, this corresponds to the set of $N$-round nonmeasuring co-strategies of Refs. 8 and 9. Note also that for any finite dimensional Hilbert space $\mathcal{L}'$,

$$\begin{aligned}\mathcal{C}_{\tilde{\mathcal{C}}_n}(\mathcal{H}_n\otimes\cdots\otimes\mathcal{H}_0, \mathcal{L}') &= \{Y \geq 0, \mathrm{Tr}_{\mathcal{L}'} Y \in \mathcal{C}_n = \mathcal{C}_{\mathcal{C}_{n-1}}(\mathcal{H}_{n-1}\otimes\cdots\otimes\mathcal{H}_0, \mathcal{H}_n)\}\\ &= \{Y \geq 0, \mathrm{Tr}_{\mathcal{H}_n}(\mathrm{Tr}_{\mathcal{L}'} Y) \in \widetilde{\mathcal{C}_{n-1}}\}\\ &= \mathcal{C}(\mathcal{H}_0, \ldots, \mathcal{H}_n\otimes\mathcal{L}').\end{aligned}$$

Now we obtain the following expressions for the corresponding norm and its dual.

**Theorem 3.** *Let $n \geq 2$. Let $X \in B_h(\mathcal{H}_n\otimes\cdots\otimes\mathcal{H}_0)$ and let $\Phi : B(\mathcal{H}_{n-1}\otimes\cdots\otimes\mathcal{H}_0) \to B(\mathcal{H}_n)$ be the corresponding map. We have*

$$\begin{aligned}\|X\|_{\mathcal{C}(\mathcal{H}_0,\ldots,\mathcal{H}_n)} &= \sup_{Y_1,Y_2\geq 0, Y_1+Y_2\in\mathcal{C}(\mathcal{H}_0,\ldots,\mathcal{H}_n,\mathbb{C})} \mathrm{Tr}\, X(Y_1 - Y_2)\\ &= \sup_{Y\in\mathcal{C}(\mathcal{H}_0,\ldots,\mathcal{H}_n,\mathbb{C})} \|Y^{1/2}XY^{1/2}\|_1\\ &= \inf_{Y\in\mathcal{C}(\mathcal{H}_0,\ldots,\mathcal{H}_n)} \inf\{\lambda > 0, -\lambda Y \leq X \leq \lambda Y\}\\ &= \sup_{\dim(\mathcal{L}')<\infty}\ \sup_{Y\in\mathcal{C}(\mathcal{H}_0,\ldots,\mathcal{H}_{n-2},\mathcal{H}_{n-1}\otimes\mathcal{L}')} \|(\Phi\otimes id_{\mathcal{L}'})(Y)\|_1\\ &= \sup_{Y\in\mathcal{C}(\mathcal{H}_0,\ldots,\mathcal{H}_{n-2},\mathcal{H}_{n-1}\otimes\mathcal{L})} \|(\Phi\otimes id_{\mathcal{L}})(Y)\|_1,\end{aligned}$$

*where $\dim(\mathcal{L}) = \dim(\mathcal{H}_{n-1}\otimes\cdots\otimes\mathcal{H}_0)$. Moreover, the dual norm is*

$$\|X\|_{I_{\mathcal{H}_n}\otimes\mathcal{C}(\mathcal{H}_0,\ldots,\mathcal{H}_{n-1})} = \|X\|_{\mathcal{C}(\mathcal{H}_0,\ldots,\mathcal{H}_n,\mathbb{C})}.$$

*Proof.* Duality of the norms is obtained from Theorem 2, this also implies the first equality. Next two equalities follow by Corollary 2. The rest follows by Theorem 2. □

For $n = 2N - 1$, first two expressions are exactly the $N\diamond$-norm as obtained in Refs. 9 and 3. Duality of the norms corresponding to strategies and co-strategies was also obtained in Ref. 9.

## IV. A GENERAL QUANTUM DECISION THEORY

As before, let $B$ be a faithful section of a base of $B(\mathcal{H})^+$. As we have seen, elements of $B$ may represent certain quantum devices and it is therefore reasonable to consider the following definitions.

Let $\{b_\theta, \theta \in \Theta\} \subset B$ be a parametrized family, for simplicity, we will suppose that the set of parameters $\Theta$ is finite. If $B$ is the set of states, the pair $\mathcal{E} = (\mathcal{H}, \{b_\theta, \theta \in \Theta\})$ is called an experiment and is interpreted as an *a priori* information on the true state of the system. Accordingly, for a section $B$, we define a generalized experiment as a triple $\mathcal{E} = (\mathcal{H}, B, \{b_\theta, \theta \in \Theta\})$.

Another ingredient of decision theory is a (finite) set $D$, the set of possible decisions. A decision procedure $\mathbf{m}$ is a procedure by which we pick some decision $d \in D$, with probability based on the "true value" of $b$. That is, $\mathbf{m}$ is a map $B \to \mathcal{P}(D)$, where $\mathcal{P}(D)$ is the set of probability measures on $D$, such a map will be called a measurement on $B$, with values in $D$. The payoff obtained if $d \in D$ is chosen while the true value is $\theta \in \Theta$ is given by the payoff function $w : \Theta \times D \to [0, 1]$, the pair $(D, w)$ is called a (classical) decision problem. Let $\lambda$ be an *a priori* probability distribution on $\Theta$. The task is to maximize the average payoff, that is the value of

$$\mathcal{L}_{\mathcal{E},\lambda,w}(\mathbf{m}) := \sum_{\theta,d} \lambda_\theta w(\theta, d)\mathbf{m}(b_\theta)_d \tag{12}$$

over all measurements $\mathbf{m} : B \to \mathcal{P}(D)$.

It is quite clear that any measurement $\mathbf{m}$ on $B$ is given by a collection $\{\mathbf{m}_d, d \in D\}$ of elements in $Q^*$ such that $\mathbf{m}(b)_d = \langle \mathbf{m}_d, b\rangle$ and that we must have $\sum_d \mathbf{m}_d = e_B$. Similarly as it was shown in Ref. 13, any measurement is given by a collection $\{M_d, d \in D\} \subset B(\mathcal{H})^+$ such that $\mathbf{m}_d = \pi(M_d)$ and $\pi(\sum_d M_d) = e_B$, that is

$$\sum_d M_d \in \pi^{-1}(e_B) \cap B(\mathcal{H})^+ = \tilde{B}.$$

Any such collection of positive operators will be called a generalized POVM (with respect to $B$), or a $B$-POVM. It is also clear that any $B$-POVM defines a measurement on $B$ (but it may happen that different generalized POVMs define the same measurement, see Ref. 13). If $B = \mathfrak{S}(\mathcal{H})$, we obtain a (usual) POVM $M = \{M_d, d \in D\} \subset B(\mathcal{H})^+$, $\sum_d M_d = I$.

Let us denote by $\mathcal{M}_B(\mathcal{H}, D)$ the set of all generalized POVMs with respect to $B$ with values in $D$ and let $\{M_d, d \in D\} \in \mathcal{M}_B(\mathcal{H}, D)$. Let us denote

$$M = \sum_{d\in D} |d\rangle\langle d| \otimes M_d^{\mathsf{T}} \in B(\mathcal{H}_D \otimes \mathcal{H})^+, \tag{13}$$

where $\mathcal{H}_D$ is a Hilbert space with $\dim(\mathcal{H}_D) = |D|$ and $\{|d\rangle, d \in D\}$ an ONB in $\mathcal{H}_D$. Then it is clear that $M$ is a block-diagonal element in $\mathcal{C}_B(\mathcal{H}, \mathcal{H}_D)$. Conversely, it is clear that if $X = \sum_d |d\rangle\langle d| \otimes X_d \in \mathcal{C}_B(\mathcal{H}, \mathcal{H}_D)$, then $\{X_d^{\mathsf{T}}, d \in D\} \in \mathcal{M}_B(\mathcal{H}, D)$. In this way, we identify $\mathcal{M}_B(\mathcal{H}, D)$ with the subset of block-diagonal elements in $\mathcal{C}_B(\mathcal{H}, \mathcal{H}_D)$.

Let now $(D, w)$ be a decision problem and let $\mathbf{m}$ be a decision procedure with corresponding $B$-POVM $M$. Then the average payoff is computed as

$$\mathcal{L}_{\mathcal{E},\lambda,w}(\mathbf{m}) = \mathcal{L}_{\mathcal{E},\lambda,w}(M) := \sum_{\theta,d} \lambda_\theta w(\theta, d)\mathrm{Tr}\, M_d b_\theta = \mathrm{Tr}\, \xi_{\mathcal{E},\lambda,w} M^{\mathsf{T}},$$

where

$$\xi_{\mathcal{E},\lambda,w} = \sum_\theta \sum_d \lambda_\theta w(\theta, d)|d\rangle\langle d| \otimes b_\theta = \sum_d |d\rangle\langle d| \otimes \bar{b}_d \in B(\mathcal{H}_D \otimes \mathcal{H})^+,$$

where $\bar{b}_d := \sum_\theta \lambda_\theta w(\theta, d) b_\theta$.

More generally, let $\mathcal{D}$ be a Hilbert space, $\dim(\mathcal{D}) = k$ and let $W$ be a function $W : \theta \mapsto W_\theta \in B(\mathcal{D})^+$, with $W_\theta \leq I$. We call the pair $(\mathcal{D}, W)$ a quantum decision problem.[15] Mathematically, this is a natural extension of classical decision problems, but at present its operational relevance is not clear.

A decision procedure is now a $B$-channel $\Phi : B(\mathcal{H}) \to B(\mathcal{D})$ and the average payoff of $\Phi$ is given by

$$\mathcal{L}_{\mathcal{E},\lambda,W}(\Phi) = \sum_\theta \lambda_\theta \mathrm{Tr}\, \Phi(b_\theta) W_\theta.$$

If $X \in \mathcal{C}_B(\mathcal{H}, \mathcal{D})$ is the Choi matrix of $\Phi$, then the average payoff has the form

$$\begin{aligned} \mathcal{L}_{\mathcal{E},\lambda,W}(\Phi) = \mathcal{L}_{\mathcal{E},\lambda,W}(X) &:= \sum_\theta \lambda_\theta \mathrm{Tr}\,(W_\theta \mathrm{Tr}_{\mathcal{H}}[(I_{\mathcal{D}} \otimes b_\theta^{\mathsf{T}})X]) \\ &= \sum_\theta \mathrm{Tr}\,(\lambda_\theta W_\theta \otimes b_\theta^{\mathsf{T}})X = \mathrm{Tr}\, \xi_{\mathcal{E},\lambda,W} X^{\mathsf{T}}, \end{aligned} \tag{14}$$

where

$$\xi_{\mathcal{E},\lambda,W} = \sum_\theta \lambda_\theta W_\theta^{\mathsf{T}} \otimes b_\theta \in B(\mathcal{D} \otimes \mathcal{H})^+.$$

It is easy to see that the set of quantum decision problems contains also classical ones: Let $(D, w)$ be a classical decision problem and let $\mathcal{H}_D$ be as before. Let $W_\theta := \sum_{d\in D} w(\theta, d)|d\rangle\langle d|$, then $(\mathcal{H}_D, W)$ is a quantum decision problem and $\xi_{\mathcal{E},\lambda,W} = \xi_{\mathcal{E},\lambda,w}$. Let $X \in \mathcal{C}_B(\mathcal{H}, \mathcal{H}_D)$ and $X = \sum_{c,\, d\in D} |c\rangle\langle d| \otimes X_{cd}$ $X_{cd} \in B(\mathcal{H})$. Since $\xi_{\mathcal{E},\lambda,w}$ is block-diagonal, we have

$$\mathcal{L}_{\mathcal{E},\lambda,W}(X) = \mathcal{L}_{\mathcal{E},\lambda,w}(M),$$

where $M = \sum_d |d\rangle\langle d| \otimes X_{dd}$ is a $B$-POVM. In other words, for a classical decision problem one cannot get better results by considering quantum decision procedures. Conversely, let $(\mathcal{D}, W)$ be a quantum decision problem such that all the operators $W_\theta$ commute. Then there is a basis of $\mathcal{D}$ with respect to which all the operators $W_\theta$ are given by diagonal matrices, and the problem is equivalent to a classical problem, in the sense that we obtain the same average payoffs. Hence we can view the set of classical decision problems as the subset of quantum decision problems such that the payoff function $W$ has commutative range.

**Theorem 4.** *Let $\mathcal{E} = (\mathcal{H}, B, \{b_\theta, \theta \in \Theta\})$ be a generalized experiment and let $(\mathcal{D}, W)$ be a quantum decision problem. Then the maximal average payoff is given by*

$$\mathcal{L}_{\mathcal{E},\lambda,W} := \max_{X\in\mathcal{C}_B(\mathcal{H},\mathcal{D})} \mathcal{L}_{\mathcal{E},\lambda,W}(X) = \|\xi_{\mathcal{E},\lambda,W}\|_{I_{\mathcal{D}}\otimes B}.$$

*If $(\mathcal{D}, W)$ is classical, then*

$$\mathcal{L}_{\mathcal{E},\lambda,W} = \inf_{b\in B} \sup_{d\in D} 2^{D_{\max}(\bar{b}_d\|b)}.$$

*Proof.* By (14), the maximal average payoff is given by

$$\mathcal{L}_{\mathcal{E},\lambda,W} = \max_{X\in\mathcal{C}_B(\mathcal{H},\mathcal{D})} \mathrm{Tr}\, \xi_{\mathcal{E},\lambda,W} X^{\mathsf{T}} = \|\xi_{\mathcal{E},\lambda,W}\|_{I_{\mathcal{D}}\otimes B},$$

the last equality follows by Corollary 3 and Proposition 1. If $(\mathcal{D}, W)$ is classical, then we may suppose that the matrices $W_\theta$ are diagonal. Then $\xi_{\mathcal{E},\lambda,W} = \sum_d |d\rangle\langle d| \otimes \bar{b}_d$ is block-diagonal. By Corollary 3, and definition of $D_{max}$,

$$\begin{aligned} \|\xi_{\mathcal{E},\lambda,W}\|_{I_{\mathcal{D}}\otimes B} &= \inf_{b\in B} 2^{D_{max}(\xi_{\mathcal{E},\lambda,W}\|I_{\mathcal{D}}\otimes b)} = \inf_{b\in B} \inf\{\gamma > 0, \bar{b}_d \leq 2^\gamma b, \forall d \in D\} \\ &= \inf_{b\in B} \sup_{d\in D} 2^{D_{max}(\bar{b}_d\|b)}. \end{aligned}$$

□

We can also use Corollary 4 to characterize decision procedures that maximize average payoff, we will call such procedures optimal with respect to $(\mathcal{E}, \lambda, W)$.

*Corollary 5. Let $(\mathcal{D}, W)$ be a decision problem and let $X \in \mathcal{C}_B(\mathcal{H}, \mathcal{D})$. Then $X$ is optimal with respect to $(\mathcal{E}, \lambda, W)$ if and only if there is some element $q \in \mathrm{span}(B)$ such that $\xi_{\mathcal{E},\lambda,W} \leq I_{\mathcal{D}} \otimes q$ and*

$$((I \otimes q) - \xi_{\mathcal{E},\lambda,W})X^{\mathsf{T}} = 0. \tag{15}$$

*If $(\mathcal{D}, W)$ is classical, then a $B$-POVM $(M_1, \ldots, M_{\dim(\mathcal{D})})$ is optimal if and only if there is some $q \in \mathrm{span}(B)$ such that $\bar{b}_d \leq q$ for all $d$ and*

$$q \sum_d M_d = \sum_d \bar{b}_d M_d. \tag{16}$$

*Proof.* The first part follows directly by Theorem 4 and Corollary 4. If $(\mathcal{D}, W)$ is classical, then $\xi_{\mathcal{E},\lambda,W}$ is block-diagonal, so that $\xi_{\mathcal{E},\lambda,W} \leq I \otimes q$ if and only if each block is majorized by $q$, that is, $\bar{b}_d \leq q$. Moreover, (16) implies that

$$\sum_d \mathrm{Tr}\,(q - \bar{b}_d)M_d = 0.$$

Since this is a sum of nonnegative elements, it is zero if and only if each summand is equal to zero. Again by positivity, this is equivalent to (15). □

In particular, in the case $B = \mathfrak{S}(\mathcal{H})$, we obtain the following optimality condition for POVMs.

*Corollary 6. Let $\mathcal{E} = \{\sigma_\theta, \theta \in \Theta\}$ be an experiment and let $(D, w)$ be a classical decision problem. Then a POVM $\{M_d, d \in D\}$ is optimal with respect to $(\mathcal{E}, \lambda, W)$ if and only if $q := \sum_d \bar{\sigma}_d M_d$ is hermitian and such that $\bar{\sigma}_\theta \leq q$ for all $d$, here $\bar{\sigma}_\theta := \sum_\theta \lambda_\theta \sigma_\theta w(\theta, d)$.*

*Remark 1.* Sometimes the function $W$ is interpreted as loss rather than payoff, then $\mathcal{L}_{\mathcal{E},\lambda,W}(\Phi)$ is the average loss of the procedure $\Phi$ which has to be minimized. Let $W'_\theta = I_{\mathcal{D}} - W_\theta$, then $\theta \mapsto W_\theta$ is again a payoff (or loss) function and we have

$$\begin{aligned}\min_\Phi \mathcal{L}_{\mathcal{E},\lambda,W} &= \min_\Phi \sum_\theta \lambda_\theta \mathrm{Tr}\,\Phi(b_\theta)W_\theta = \min_\Phi \sum_\theta \lambda_\theta \mathrm{Tr}\,\Phi(b_\theta)(I - W'_\theta)\\ &= 1 - \max_\Phi \mathcal{L}_{\mathcal{E},\lambda,W'}(\Phi) = 1 - \|\xi_{\mathcal{E},\lambda,W'}\|_{I_{\mathcal{D}}\otimes B}.\end{aligned}$$

Moreover, an optimal procedure $\Phi$ that minimizes the loss is a maximizer for $\mathcal{L}_{\mathcal{E},\lambda,W'}$, hence satisfies the conditions of Corollary 5, with $W$ replaced by $W'$. Note that then the condition from Corollary 6 is the same as obtained in Ref. 12.

Let $\{M_d, d \in D\}$ be a $B$-POVM with $\sum_d M_d = c \in \tilde{B}$. Then since $0 \leq M_d \leq c$ for all $d$, we have

$$M_d = c^{1/2}\Lambda_d c^{1/2}, \qquad d \in D,$$

where $\Lambda_d := c^{-1/2}M_d c^{-1/2}$ defines a (usual) POVM on the support $\mathrm{supp}\,c$ of $c$. It follows that $\mathrm{Tr}\,xM_d = \mathrm{Tr}\,c^{1/2}xc^{1/2}\Lambda_d$, that is, we can decompose the measurement defined by $\{M_d\}$ into a cp map $\chi_c$: $x \mapsto c^{1/2}xc^{1/2}$ followed by the usual measurement given by $\{\Lambda_d\}$. Note that $\chi_c \in \mathcal{C}_B(\mathcal{H}, \mathrm{supp}\,c)$ so that $\chi_c$ maps a generalized experiment $\mathcal{E} = (\mathcal{H}, B, \{b_\theta, \theta \in \Theta\})$ onto an ordinary experiment $\mathcal{E}_c := \{\mathrm{supp}\,c, \mathfrak{S}(\mathrm{supp}\,c), \{\chi_c(b_\theta), \theta \in \Theta\}\}$. We write this decomposition as $M = \Lambda \circ \chi_c$. Such a decomposition was also used in Ref. 3 in the case of testers and in Ref. 13 for generalized POVMs. Using this decomposition, we obtain the following optimality condition for $B$-POVMs.

*Corollary 7. Let $(D, w)$ be a classical decision problem and let $M \in \mathcal{M}_B(\mathcal{H}, D)$ with decomposition $M = \Lambda \circ \chi_c$. Suppose $c$ is invertible and let $\mathcal{E}_c := (\mathcal{H}, \{\sigma_\theta := \chi_c(b_\theta), \theta \in \Theta\})$. Then $M$ is optimal for $(\mathcal{E}, \lambda, w)$ if and only if $\Lambda$ is optimal for $(\mathcal{E}_c, \lambda, w)$ and*

$$\sum_d \bar{\sigma}_d \Lambda_d \in \mathrm{span}(\chi_c(B)),$$

*where $\bar{\sigma}_d = \sum_\theta \lambda_\theta w(\theta, d)\sigma_\theta$.*

*Proof.* Directly by Corollaries 5 and 6. □

*Example 2 (Multiple hypothesis testing).* Suppose a family $\{b_1, \ldots, b_k\}$ of elements in $B$ is given and the task is to decide which is the true one, moreover, given some $\lambda \in \mathcal{P}(\{1, \ldots k\})$, we want to minimize the average probability of making an error. In this case, put $\mathcal{E} = (\mathcal{H}, B, \{b_1, \ldots, b_k\})$, $\Theta = D = \{1, \ldots, k\}$ and the loss function is $w(i, j) = 1 - \delta_{ij}$, where $\delta$ is the Kronecker symbol. A decision procedure is a $B$-POVM $\{M_1, \ldots, M_k\}$, where $M_i$ corresponds to the choice $b_i$. Then the average loss is the average error probability

$$\mathcal{L}_{\mathcal{E},\lambda,w}(M) = \sum_{i,j} \lambda_i (1 - \delta_{ij}) \mathrm{Tr}\, b_i M_j = \sum_{i \neq j} \lambda_i \mathrm{Tr}\, b_i M_j.$$

We can use Remark 1 to compute the minimal average error probability $\Pi_\lambda^B(b_1, \ldots, b_k) := \min_M \mathcal{L}_{\mathcal{E},\lambda,w}(M)$. We obtain $\xi_{\mathcal{E},\lambda,w'} = \sum_i |i\rangle\langle i| \otimes \lambda_i b_i$, so that the minimal average error probability is

$$\Pi_\lambda^B(b_1, \ldots, b_k) = 1 - \|\xi_{\lambda,w'}\|_{I\otimes B} = 1 - \inf_{b\in B} \sup_{1\leq i\leq k} 2^{D_{max}(\lambda_i b_i \| b)}.$$

For $B = \mathfrak{S}(\mathcal{H})$, the last equality was obtained in Ref. 7, see also Ref. 17.

Let us now look at an optimal decision procedure. Let $\{M_i\}$ be a $B$-POVM with decomposition $M = \Lambda \circ \chi_c$ and let us suppose that $c = \sum_i M_i$ is strictly positive. Let $\sigma_i = \chi_c(b_i)$ and $\mathcal{E}_c = (\mathcal{H}, \mathfrak{S}(\mathcal{H}), \{\sigma_1, \ldots, \sigma_k\})$. Suppose that $\{\Lambda_i\}$ is optimal for $(\mathcal{E}_c, \lambda, w)$, this is equivalent to the fact that $\sum_i \lambda_i \sigma_i \Lambda_i =: p$ is a hermitian element that majorizes $\lambda_i \sigma_i$ for all $i$. By Remark 1 and Corollary 7, $\{M_i\}$ is then optimal for $(\mathcal{E}, \lambda, w)$ if and only if $p \in \mathrm{span}(\chi_c(B))$, note that $\sigma_i \in \chi_c(B)$ for all $i$.

*Example 3 (Hypothesis testing).* Let $k = 2$ in the previous example, then we obtain the hypothesis testing or discrimination problem, considered at the end of Sec. I B. Here we have

$$\| |0\rangle\langle 0| \otimes s b_0 + |1\rangle\langle 1| \otimes t b_1 \|_{I_2 \otimes B} = \frac{1}{2}(\|s b_0 - t b_1\|_B + s + t),$$

for $s, t > 0$, so that indeed, $1 - \|\xi_{\mathcal{E},\lambda,w'}\|_{I_2\otimes B} = \frac{1}{2}(1 - \|\lambda b_0 - (1-\lambda) b_1\|_B)$ is the minimal Bayes error probability. Let $\{M_0, M_1\}$ be a $B$-POVM such that $c = M_0 + M_1$ is strictly positive and let $\sigma_i = \chi_c(b_i)$. Suppose $\lambda = 1/2$ and let $\Lambda_i = c^{-1/2} M_i c^{-1/2}$ be a POVM which is optimal for $(\mathcal{E}_c, \lambda, w)$, then $\Lambda_0$ is the projection onto the support of $(\sigma_0 - \sigma_1)_+$ and $\sum_i \lambda_i \sigma_i \Lambda_i = \frac{1}{2}((\sigma_0 - \sigma_1)_+ + \sigma_1)$. From the previous example, it is clear that $\{M_0, M_1\}$ is then an optimal test for $(\mathcal{E}, \lambda, w)$ if and only if any of (and therefore all of) $(\sigma_0 - \sigma_1)_+, (\sigma_0 - \sigma_1)_-, |\sigma_0 - \sigma_1|$ is an element in $\mathrm{span}(\chi_c(B))$.

In particular, let $B = \mathcal{C}(\mathcal{H}, \mathcal{K})$. In this case, the $B$-POVMs are exactly the quantum 1-testers of Refs. 3 and 21, see also Ref. 13. More precisely, the $B$-POVMs $M = \{M_d, d \in D\} \subset B(\mathcal{K} \otimes \mathcal{H})^+$ satisfy $\sum_d M_d = I \otimes \sigma$ for some $\sigma \in \mathfrak{S}(\mathcal{H})$. Let $M = \Lambda \circ \chi_{I\otimes\sigma}$ be the decomposition of $M$, then for $X_\Phi$,

$$\mathrm{Tr}\, M_d X_\Phi = \mathrm{Tr}\, \Lambda_d \chi_{I\otimes\sigma}(X_\Phi) = \mathrm{Tr}\, \Lambda_d (\Phi \otimes id_A)(\rho),$$

where $\rho = \chi_{I\otimes\sigma}(\Psi)$ is a pure state in $\mathfrak{S}(\mathcal{H} \otimes \mathcal{H}_A)$ and $\mathcal{H}_A = \mathrm{supp}\,(\sigma)$. This means that the tester $M$ is implemented by the triple $(\mathcal{H}_A, \rho, \Lambda)$. If $\sigma = \dim(\mathcal{H})^{-1} I$, then $\rho = \dim(\mathcal{H})^{-1}\Psi$ is the maximally entangled state in $\mathcal{H} \otimes \mathcal{H}$. By the results of Example 3, we have the following.

*Corollary 8. Let $b_i = X_{\Phi_i}$ be Choi matrices of the channels $\Phi_0, \Phi_1 : B(\mathcal{H}) \to B(\mathcal{K})$. Consider the problem of testing the hypothesis $\Phi_0$ against $\Phi_1$, with a priori probability $\lambda \in [0, 1]$. Then there exists an optimal 1-tester implemented by a triple $(\mathcal{H}, \Lambda, \rho)$ with maximally entangled input state $\rho$ if and only if $\mathrm{Tr}_{\mathcal{K}} |\lambda X_{\Phi_0} - (1-\lambda) X_{\Phi_1}|$ is a multiple of $I_{\mathcal{H}}$.*

## ACKNOWLEDGMENTS

This work was supported by the Grant No. VEGA 2/0125/13 and by Science and Technology Assistance Agency under the Contract No. APVV-0178-11.

[1] M. D. Choi, "Completely positive maps on complex matrices," Lin. Alg. Appl. **10**, 285–290 (1975).
[2] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Quantum circuit architecture," Phys. Rev. Lett. **101**, 060401 (2008).
[3] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Memory effects in quantum channel discrimination," Phys. Rev. Lett. **101**, 180501 (2008).
[4] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Theoretical framework for quantum networks," Phys. Rev. A **80**, 022339 (2009).
[5] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Transforming quantum operations: Quantum supermaps," Europhys. Lett. **83**, 30004 (2008).
[6] N. Datta, "Min- and max- relative entropies and a new entanglement monotone," IEEE Trans. Inf. Theory **55**, 2816–2826 (2009).
[7] N. Datta and M. Mosonyi, "Generalized relative entropies and the capacity of classical-quantum channels," J. Math. Phys. **50**, 072104 (2009).
[8] G. Gutoski and J. Watrous, "Toward a general theory of quantum games," in *Proceedings of the 39th ACM Symposium on Theory of Computing (STOC 2007)* (ACM, 2007), pp. 565–574; e-print arXiv:quant-ph/0611234v2.
[9] G. Gutoski, "On a measure of distance for quantum strategies," J. Math. Phys. **53**, 032202 (2012).
[10] C. E. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
[11] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North Holland, Amsterdam, 1982).
[12] A. Holevo, *Statistical Structure of Quantum Theory* (Springer-Verlag, Berlin, 2001).
[13] A. Jenčová, "Generalized channels: channels for convex subsets of the state space," J. Math. Phys. **53**, 012201 (2012); e-print arxiv:1105.1899.
[14] A. Kitaev, "Quantum computations: algorithms and error correction," Russian Math. Surv. **52**, 1191–1249 (1997).
[15] K. Matsumoto, "A quantum version of randomization criterion," preprint arxiv:1012.2650v6 (2012).
[16] V. Paulsen, *Completely Bounded Maps and Operator Algebras* (Cambridge University Press, 2003).
[17] R. Koenig, R. Renner, and C. Schaffner, "The operational meaning of min- and max-entropy," IEEE Trans. Inf. Theory **55**, 4337 (2009).
[18] T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1970).
[19] J. Watrous, "Notes on super-operator norms induced by Schatten norms," Quantum Inf. Comput. **5**, 58–68 (2005); e-print arxiv:quant-ph/0411077v1.
[20] D. Reeb, M. J. Kastoryano, and M. M. Wolf, "Hilbert's projective metric in quantum information theory," J. Math. Phys. **52**, 082201 (2011).
[21] M. Ziman, "Process POVM: A mathematical framework for the description of process tomography experiments," Phys. Rev. A **77**, 062112 (2008).